METAMEMORY IN MULTIPLE SCLEROSIS.

A thesis submitted for the degree of Doctor of Philosophy

by

Austin M Claffey

School of Health Sciences & Social Care

Brunel University

May 2010.

Abstract

The concept of metamemory proposes that supplementary to typically measured memory abilities, memory monitoring and control processes are used to optimise learning. Accurate memory monitoring appears to be underpinned by a range of cognitive, and possibly affective, contributions. In populations with these deficits, metamemory has been shown to be impaired. In Multiple Sclerosis (MS), only a limited metamemory literature exists, surprising given that MS is a leading cause of disability among people of working age, and cognitive and mood disorder is common.

Using structural equation modelling, this study of 100 people with MS explored factors contributing to performance on episodic Judgment of Learning, Retrospective Confidence and Feeling of Knowing. Given its negative influence on cognitive domains in MS, the impact of information processing deficits on metamemory was also investigated. Finally, memory self-report, a frequently used clinical indicator of memory functioning, was assessed.

Findings suggest that memory complaint is associated with mood, and is unrelated to tested memory. Second, Retrospective Confidence Judgments were predictive of memory performance, even in the presence of memory impairment. Third, an unusual finding of maintained underconfidence at delay was observed in the Judgment of Learning task. Finally, Feeling of Knowing judgments related to executive, but not to memory ability. A novel finding in respect of this judgment was of processing speed relating negatively to accuracy, in the context of executive dysfunction. This suggests that some top-down direction of processing resources may be a factor in supporting accuracy, rather than the speed at which information is processed. Of all the task-based judgments, accuracy in this judgment was the only one with a reliable association with mood. Faster processing speed, executive dysfunction and least depression symptomatology related to low accuracy, perhaps typifying a profile of disinhibition seen in MS, characterised by poorly constrained processing and apparently elevated mood.

Table of Contents

ACKNOWLEDGEMENTS

Two people, Dr Priscilla Harries and Professor Lorraine DeSouza, who have sat with me in many rooms throughout this process, deserve serious gratitude. I have likened them in my own mind to my global and local processors, combining attention to detail and vision. With them, and Dr Stanley Gaines, this has been a wonderful learning experience. For your time, advice, availability, support and encouragement, I want to say 'Thank You.'

I have been so impressed by the interest shown in this research by charities, support organisations, and the MS grapevine. Most of all, gratitude must go to the people who participated, advertised, even cooked for me, in a combined effort to support the study. From them I have learned much about the importance of attitude in managing this disease. I hope this study does them justice.

Friends and family have been at the sharp end of me during the past few years; specifically I would like to acknowledge Teresa Bergin, Christine & David McKinney, Maggie Burrows, Mo Norris, Cathryn Edmonds & 'the Claffeys'.

Finally, getting a paragraph all of his own, I would like to thank Colin Kenny. Throughout he has been an unswerving support, facilitating every step of the process. Especially during the many long days of data collection and more recently in writing, I have been thoroughly sustained by him.

Figures & Formulae

Tables

Page

Appendices

A. Sentence Memory Test materials
B. Ethical Approval for the Study
C. Participant Information Sheet
D. CFA Model 1: Beck Depression Inventory 1 Factor
E. CFA Model 2: Beck Depression Inventory 2 Factor (i)
F. CFA Model 3: Beck Depression Inventory 2 Factor (ii)
G. CFA Model 4: Beck Depression Inventory 3 Factor
H. Confirmatory Factor Analysis: a priori Cognitive Model 3-Factor
I. Confirmatory Factor Analysis: Cognitive Model 4-Factor
J. Confirmatory Factor Analysis: General Frequency of Forgetting 1 Factor
K. Confirmatory Factor Analysis: General Frequency of Forgetting 3 Factor
L. Structural Model: General Frequency of Forgetting
M. Structural Model: Forgetting While Reading
N. Structural Model: Judgment of Learning direct effects
O. Structural Model: Judgment of Learning direct effects, including Information Processing
P. Structural Model: Retrospective Confidence - Absolute Accuracy
Q. Structural Model: Retrospective Confidence - Relative Accuracy
R. Structural Model: Feeling of Knowing - Absolute Accuracy
S. Structural Model: Feeling of Knowing - Inaccuracy Mediation
T. Structural Model: Feeling of Knowing - Relative Accuracy
U. Structural Model: Feeling of Knowing - Relative Accuracy, including Information Processing
V. Structural Model: Feeling of Knowing - Mediation Model

Structure of the Thesis.

In Chapter 1 *Literature Review,* the context of the study is presented. In people with neurological illness, the relationships between neuropsychological impairment and learning are considered. Learning ability, especially when self-regulated, is both a method and an aim in neurological rehabilitation. A review of one common neurological disease, Multiple Sclerosis (MS), focusing on neuropsychological and affective consequences is then presented. The aim is to identify the key areas of deficit potentially impacting on learning ability. The relationships between cognitive and affective deficits in MS are then considered in respect of awareness and management of cognitive abilities. Specifically, the review focuses on the accurate monitoring of memory abilities, or metamemory. Accurate monitoring of memory is proposed to be a key component of learning, because it facilitates appropriate learning-oriented behaviours (Prigatano, 1999).

The theoretical underpinning of metamemory is then reviewed, focusing on the mechanisms for how a range of monitoring decisions might be made; affect-based, using memory-experiences or using inference. The measurement of metamemory from the point of view of self-efficacy, or subjective report of memory is first considered. A second measurement approach focusing on accuracy drawn from specific memory tasks, or 'on-line' metamemory judgment (Toglia & Kirk, 2000) is reviewed. Both approaches are addressed because clinical practice may orient towards the former, and research literature the latter.

The chapter ends by drawing together the literature on metamemory in neurological populations, then to metamemory in Multiple Sclerosis. From the review, a number of objectives are developed, from which the study then proceeds by proposing a number of a priori models of cognitive and affective contributions to four metamemory judgments. These latent variable models investigate factors that are proposed to contribute to accuracy in metamnemonic judgments.

In Chapter 2, *Development of Methods,* the selection of relevant measures for proposed cognitive, affective and metamemory variables for the study is considered, based on psychometric properties of instruments and on their appropriateness to the sample. Chapter 3, *Development of Statistical Methods,* presents a review of the selected method of analysis, structural equation modelling. The chapter outlines the two steps involved in the approach. First, the investigation of each proposed latent variable using confirmatory factor analysis is outlined; this is called the measurement model because it reflects the testing of assumptions about selected tests' factorial coherence. The second step, termed the structural model, which is used to investigate relationships between latent variables, is summarised.

Chapter 4, *Methods*, outlines the major ethical issues in the study, the recruitment of the sample, and the procedures involved in collecting the data for the study. This chapter also considers statistical methods relating to data screening, transformation and missing data handling.

There are three results chapters, each of which will be presented with an initial discussion of findings. Chapter 5 *Sample Performance* outlines the demographics of the sample and performance on the range of measures used. The aims in analysing this data are, in part, to assess the features of the sample against known samples of community dwelling people with MS in the UK, so an assessment of generalisability can be made. This chapter also presents the results on the range of measures used, which will contribute to the assessment of subsequent results.

Chapter 6, *The Measurement Models* presents the results of confirmatory factor analyses for each of the proposed latent variables to be used in final modelling, relating this to a discussion of both statistical and conceptual issues raised. The results of this analysis will contribute to the final set of results; presented in Chapter 7, *The Structural Models*. This chapter will present results of testing each of the *a priori* models of subjective memory appraisal, retrospective confidence, judgment of learning and feeling of knowing. There are two components to these models' results - identifying the latent contributions for each of the metamnemonic judgments for the sample, and a consideration of how these contributions fit with proposed models of these metamemory judgments.

Chapter 8, *Discussion,* will address the objectives of the study by drawing together the results with relevant published evidence - performance data on individual measures, factorial structure of tests used in this and other samples, and performance of the sample on metamemory tasks in respect of other neurological and non-neurological samples.

Clinical implications of the study's findings will be explored and implications for further research to answer remaining, or emerging questions, will be addressed. These conclusions will also consider issues of measurement and analysis. The overall structure of the study is outlined in figure 1.1.

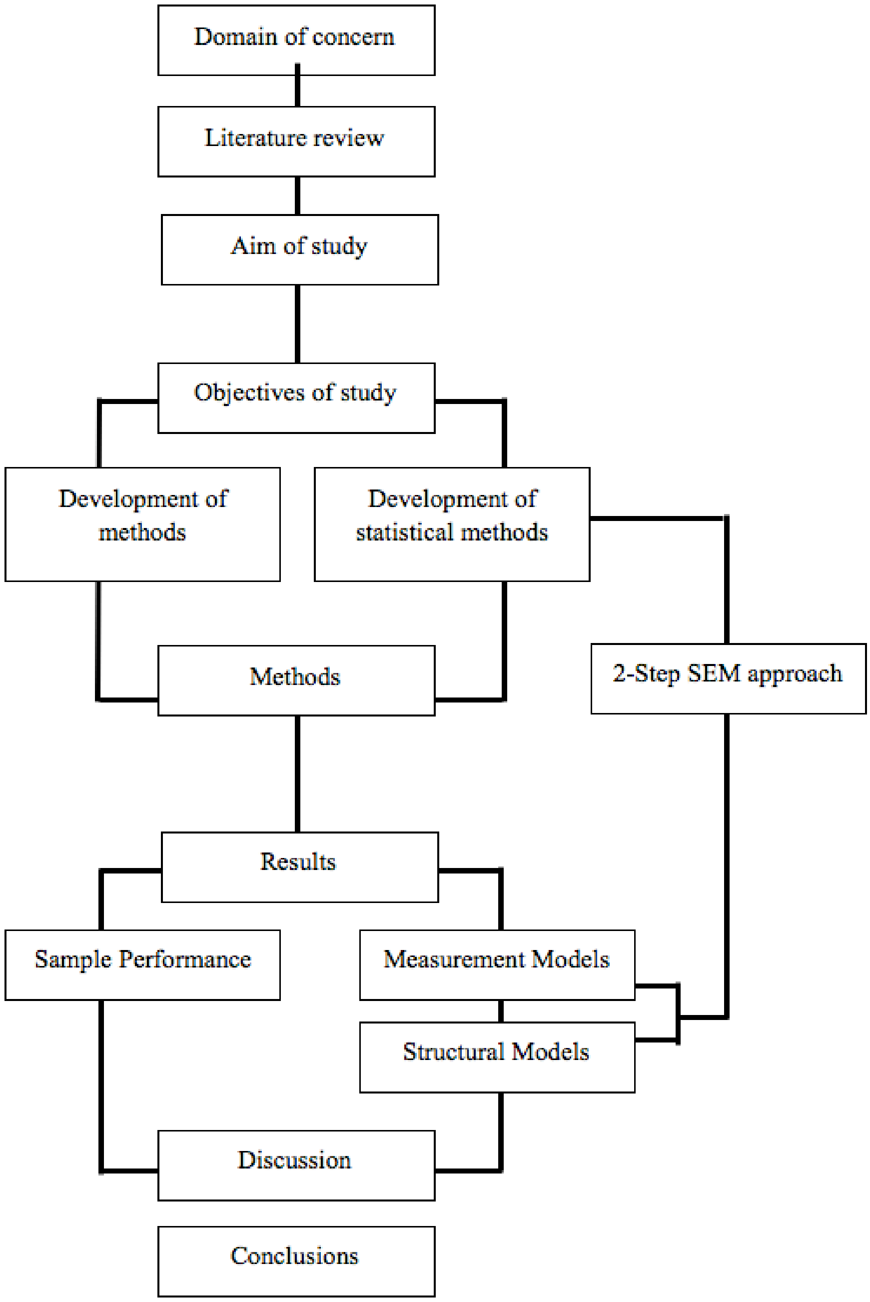


Figure 1.1. Structure of the thesis; SEM = Structural Equation Modelling

Chapter 1: Literature Review

This chapter focuses on the clinical context for the study; the repercussions of neuropsychological impairment for both learning and rehabilitation success for people with Multiple Sclerosis (MS). The introduction lays out some of the issues relating to how requisite learning ability is perhaps assumed in rehabilitation, rather than specifically planned for. The relevance of this to MS patient groups is considered. An outline of MS is then presented considering the pathology, diagnosis, symptoms and treatment of the disease. The main focus for the review of both symptoms and treatment relates to neuropsychological impairment. The process of rehabilitation is considered in respect of approaches that consider adequate awareness, monitoring and control of learning and memory, leading to a consideration of the topic of metamemory. The metamemory literature is reviewed to outline key concepts, models and methods of study, then considered in relation to performance of neurologically impaired groups in this domain. Finally, studies of metamemory in MS are appraised in order to derive the aims for the study.

1.1 Introduction.

The development of National Service Frameworks (NSF) by the United Kingdom Department of Health over the last decade were an attempt to improve healthcare service delivery, with different frameworks addressing topics such as older people, mental health and long term conditions (Great Britain. Department of Health, 1999, 2001, 2005). The NSF for long-term conditions, whose target audience ranged from NHS Trust chief executives to voluntary organisations, sets out a number of ‘quality requirements’ to best support people with long-term, mainly neurological, conditions (Great Britain. Department of Health, 2005:3).

Two key areas emphasised in the NSF and related documents are symptom self-management and the provision of specialist rehabilitation (Great Britain. Department of Health, 2005). About half of the NSF quality requirements focus on rehabilitation interventions and disease self-management - ranging from early interventions to vocationally-focused and community-based rehabilitation, all with aim of supporting independent living (Department of Health, 2005:5). Self-management was supported through the development of Expert Patient Programmes (Great Britain. Department of

Health, 2001) by which the experience-derived knowledge of people in managing chronic diseases is used as a resource for other people with chronic diseases.

Despite some concerns that Expert Patient Programmes do not reach those who need it most, an analysis from the UK suggested that, in part, it achieved a key aim in acknowledging the experience of living with chronic illness, providing some empowerment to patients (Wilson, Kendall & Brooks, 2007; Lorig et al., 2005), though additional benefits in disease self-management have not been consistently found (Rogers et al., 2008).

One limitation of such programmes appears to be that those who benefit most, as might also be the case in rehabilitation more generally, are ‘systematically organised’ people who already have intent towards learning, autonomy or ‘involvement’ (Wilson, Kendall & Brooks, 2007:430). These are frequently those who also benefit from rehabilitation based on working assumptions that learning processes are the same for people with neurological disease, including those with cognitive impairment (Prigatano, 1999; Glisky & Glisky, 2002). Patient groups with this orientation, or ability, are likely to be in possession of generally intact neuropsychological functions or lacking the negative impact of mood disorder, maladjustment or reduced awareness (Prigatano, 1999).

Links between pathology and learning impairments are often not considered adequately in relation to key learning-related factors such as speed of learning, motivation or generalisability of learning (Prigatano, 1999; Toglia & Kirk, 2000; Krakauer, 2006). One reason for the failure to adequately consider learning propensity fully may be that models of motor skill acquisition are more dominant in rehabilitation, perhaps reflecting the availability of methods of identifying functional plasticity in this domain (Winstein, Wing & Whitall, 2003). As such, a focus on repetition and performance of specific task components is often maintained, rather than one that focuses on abilities to self-regulate learning and develop task competency.

Capacity to learn is a fundamental requirement of rehabilitation, and the capacity to learn itself is built on a number of factors - motivational, cognitive, affective, personal and environmental (Wade, 1997). More specifically, in considering people with neurological illness affecting the central nervous system, mood disorder and neuropsychological functions are key intrinsic factors in learning ability (Diamond et al., 1996; Asikainen, Kaste & Sarna, 1998; Ownsworth & McKenna, 2004). Rehabilitation interventions often assume that patients are, or are capable of, ‘monitoring their understanding’ as a *good* learner does (Perfect & Schwartz, 2002:7; Fiszdon, et al., 2006). Because learning is often inadequately considered in the design of rehabilitation interventions, those who learn in a manner not impacted greatly by disease process may benefit when routine, intuitive assumptions about how learning occurs, is applied and is maintained (Galski, et al., 1993; Wilson, 1997; Cicerone et al., 2000). These assumptions are often seen in action in clinical practice where, for example, repetition blocks of a task component are carried out, with the supposition that this leads to generalisable learning (Toglia, 1991; Hanlon, 1996; Lincoln, et al., 2002). The efficacy of this approach is often supported in the research base by screening out people with neuropsychological impairment or those who cannot engage in goal-setting, itself potentially a surrogate for cognitive abilities or requisite awareness of deficits (Toglia & Kirk, 2000; Fischer, Gauggel & Trexler, 2004).

‘A theory of rehabilitation without a model of learning is a vehicle without an engine’
(Baddeley, 1993:235).

In rehabilitation contexts, the identification of potential for learning, the ability to generalise new learning, and the appraisal that successful learning has occurred, are key factors (Toglia, 1991, 1998a; Fiszdon et al., 2006). Learning is different from performance, the latter, perhaps misguidedly, sometimes the focus of rehabilitation efforts (Winstein, Wing & Whitall, 2003). The distinction between the two rests on the persistence, permanence and transfer of a ‘skill’ to different tasks or environments (Hanlon, 1996; Toglia, 1998b; Winstein, Wing & Whitall, 2003); improved performance may achieve none of these.

Learners are those who show benefit from training of some kind, improving or refining performance over repeated exposure to a task, and generalising this to different contexts (Toglia, 1998b). A complication in some domains of rehabilitation, such as in neurological populations, is that the learning process itself may be compromised by neuropsychological deficit, reduced awareness or the individuals beliefs about how the disease impacts ability (Pohl, Winstein & Onla-Or, 1997; Prigatano, 1999; Toglia & Kirk, 2000; Boyd, et al., 2007). This means that in considering learning and its application, the impairments caused by a neurological illness are important. The extent and type of neuropsychological dysfunction, mood disorder, the impact of reaction to the disease (including implicit beliefs about ability) and motivation may all be relevant factors. Each may have an impact in creating an evaluative bias. A potential result is that learning behaviours based on such distorted evaluations, are themselves inappropriate (Dunlosky & Metcalfe, 2009).

The clinical entity of Multiple Sclerosis (MS) is one such neurological illness; neuropsychological, motor, sensory and affective disturbance is common (Compston, et al., 2006). In the author’s clinical experience, management of sensory and motor disabilities caused by the illness is often emphasised at the expense of the impact of neuropsychological impairment, the latter potentially having a greater impact on learning ability. This imbalance is evident in the MS rehabilitation research base (O’Brien, et al., 2008), and in proposals that this cognitive impairment, along with fatigue and low mood, are often neglected aspects of the illness (Minden, 2000; Thomas, et al., 2006). Studies focusing on rehabilitation efficacy, where anything more than mild cognitive impairment can be a criterion for exclusion perhaps underscore this bias (Patti et al., 2002; Khan, et al., 2008). At the same time, MS-focused Cochrane reviews of the effectiveness of vocational rehabilitation (Khan, Ng & Turner-Stokes, 2009), psychological therapies (Thomas, et al., 2006), occupational therapy (Steultjens et al., 2003) and multidisciplinary rehabilitation (Khan, et al., 2008) underline the negative impact of cognitive deficits in MS on rehabilitation outcome.

Generally, cognitive impairment in MS can present as a somewhat hidden disability, in comparison to the sensory and motor impairments associated with the disease. The deficits in learning associated with cognitive impairment often become obvious only when increasing learning demands are made on patients in the rehabilitation, vocational or home environments (Khan, Ng & Turner-Stokes, 2009). Suggestions for further studies in the field attest to its importance, and perhaps, neglect.

It may be that there is an imbalance in consideration of two key factors in rehabilitation - *what* needs treatment and *how* is it to be treated (Wilson, 1997, 2002). Neuropsychological impairment may confer additional complexity to the *how* of treatment; a distortion in the learning processes itself. While there is a focus on generating high-quality interventional studies of rehabilitation efficacy, there also remains a need to examine the underlying issue of learning processes themselves, how they are impaired, and how interventions could be tailored as a result. Inclusion of people who do have limiting cognitive impairment would be one step; focusing on why treatment might fail, as a way of calibrating how the intervention is delivered, another.

A common experience, for example, in working with people with memory disorders is the failure of an external aid, such as a diary, to compensate for difficulties. Problems typically arise when there are differing perceptions (e.g. between patient and therapist) regarding memory ability (Prigatano, 1999), or where there are other neuropsychological deficits at the root of memory complaint or failure, meaning that effective diary use fails. A third element is where there are in fact no significant memory problems; instead a poorly investigated complaint of memory disorder, which actually signposts a mood disorder (Middleton et. al., 2006; Marrie et. al., 2005).

A poorly delivered intervention in this situation might reflect not accounting for problems with the accurate appraisal of memory ability, mood disorder, difficulties with the complexity of the task, or a mix of all three. It is common clinical problems such as these, with people with Multiple Sclerosis, which the author is keen to investigate further. Interventions that do not work may not be ineffective interventions, just poorly directed.

Rehabilitation efforts aimed at optimising learning typically benefit from being supported by interventions to improve patients' accurate appraisal and management of their memory (Prigatano, 1999). Awareness of memory difficulties may be positively related to outcomes of cognitive rehabilitation in Alzheimer’s disease (Clare et al., 2004) and may be an important basis for instituting compensatory strategies, which reduce disability (Wilson & Moffat, 1992). An analogue is the importance of accurately monitored learning, directing study approaches in educational settings (Hacker, et al., 2000; Hacker, Bol & Keener, 2008)

One area, which focuses on this appraisal and management of cognition generally, is that of metacognition (Perfect & Schwartz, 2002; Dunlosky & Metcalfe, 2009). In part borne of educational contexts, metacognition emphasises the monitoring and management of one’s own ongoing cognitive performance, including learning (Flavell, 1979; Dunlosky & Metcalfe, 2009). It is of interest in this study, because of its importance to the consideration of self-directed learning. Its basis in subjective reflection, offers the potential for approaches that manipulate those subjective judgments, optimising resulting cognitive behaviours. Poor learners, in effect, could become better learners, not just by improving on specific cognitive demands such as recall, but by optimising judgments about recall, that lead to effective memory-related strategies and behaviours (Dunlosky & Metcalfe, 2009; Metcalfe & Finn, 2008).

Given the strong relationship between memory and learning, and the occurrence of complaints about memory, as well as frank memory deficits in people with MS, the focus of this study will be on metamemory; the monitoring and management of memory processes (Flavell, 1979; Nelson & Narens, 1990; Dunlosky & Bjork, 2008a). Beatty (2004) suggests that the memory deficits in MS have good potential for rehabilitation in large proportions of those who have them, and an accurate appraisal of memory (metamemory) may be an important first step (Moulin, 2002).

To this end, memory monitoring in MS has been investigated in a small number of studies (Beatty & Monson, 1991; Scarrabelotti & Carroll, 1998; 1999; Randolph, Arnett & Higginson, 2001; Randolph Arnett & Freske, 2004; Phillips & Stuifbergen, 2006; Julian Merluzzi & Mohr, 2007). Of these, two specifically focus on the factors associated with accuracy of appraisal (Beatty & Monson, 1991; Randolph Arnett & Freske, 2004). The implication of findings for rehabilitation planning, is that clinicians need to understand that inaccurate appraisal of memory performance (e.g. the inaccurate report of a ‘poor’ or ‘excellent’ memory) may be equally indicative of a cognitive deficit (in memory or other domain), an affective problem, or both. The active role of the rememberer, using ‘complex evaluative and decisional processes’ (Koriat, Goldsmith & Pansky, 2000:487) is also emphasised in this approach to memory performance, implicating factors that bias or attenuate evaluations, such as congruency with mood (Barclay, et al., 1991; Gotlib, Roberts & Gilboa, 1996) or executive abilities.

In the author’s clinical experience, a specific and significant issue in managing cognitive deterioration (and its impact on everyday memory) is difficulty with monitoring performance in memory situations. Such difficulties in appraising the difficulty of to-be remembered information is one part of metamemory - monitoring processes - the other relates to the use of mechanisms to maximise functioning once an accurate monitoring judgment has been made, so-called control processes (Nelson et al., 1999). This clinical experience accords with an increasing interest in metacognitive abilities in people with Alzheimer’s Disease, (Moulin 2002) in normal ageing (Herzog, 2002; Souchay et al., 2007; Dunlosky and Metcalfe 2009) and in people with various forms of brain damage (Shimamura and Squire 1986; Pannu and Kaszniak 2005).

Mindful of the importance of learning for rehabilitation success, and the impact of learners' beliefs, attitude, mood and neuropsychological status on learning oriented behaviours (Toglia & Kirk, 2000), this study aims to extend previous work on metamemory in people with MS. It will investigate the factors that contribute to accuracy across range of metamemory monitoring judgments and consider how findings concord with models of metamemory. Finally, it will consider how findings might specify rehabilitation options in this population.

1.2 Multiple Sclerosis

Multiple Sclerosis, an autoimmune disorder, is typically characterised by inflammatory demyelination, axonal loss and gliosis (El-Moslimany & Lublin, 2008). Demyelination, in the context of MS, involves the destruction of the fatty (lipid) wrapping around axons in the central nervous system (CNS; brain & spinal cord) through an inflammatory process. This fatty wrapping, called myelin, provides the 'white' in 'white matter' and functions to insulate and facilitate neural transmission in the CNS (Kandel, Schwartz & Jessell, 2000). Underlying axonal loss has been shown to occur also, probably secondary to the myelin loss (Bjartmar, Wujek & Trapp, 2003). A final pathological event, *gliosis*, is associated with plaque formation - where glial (CNS helper) cells form fibrous 'scars' after the neurodegenerative phase of demyelination and axonal damage has taken place (Coyle, 2006). The development of lesions is common along the optic nerve, in white matter surrounding the ventricles of the brain, the brain stem, cerebellum and white matter of the spinal cord (Noseworthy et al., 2000). More recently, there is evidence that grey matter (neuronal cell bodies) is also affected by Multiple Sclerosis (Polman et al., 2006; Sanfilipo, et al., 2006).

Multiple Sclerosis has a wide range of clinical expressions, depending both on the location of degenerative lesions, as well as the type of clinical course (Smith, Samkoff & Scheinberg, 1993; El-Moslimany & Lublin, 2008). A number of different clinical courses have been described: relapsing-remitting (RRMS), primary-progressive (PPMS), secondary-progressive (SPMS), progressive-relapsing and benign; the latter two being less often used (National Collaborating Centre for Chronic Conditions, 2004; El-Moslimany &

Lublin, 2008). Each of the three main subtypes has an associated clinical pattern; episodic worsening with remission (relapsing-remitting MS), progressive deterioration in function (primary-progressive MS), or a combination of the two courses (secondary-progressive MS; El-Moslimany & Lublin, 2008).

Among the different courses of progression associated with the disease, the relapsing-remitting type is the most common, comprising 75-80% of consecutive MS referrals in some studies (Oshinsky, Elfont & Lublin, 1998; Tullman et al., 2004). In general, about 20% of all MS diagnoses are primarily progressive in that relapses are not a feature (Ebers, 2001; Compston & Coles, 2002). Most people with relapsing-remitting MS will eventually go on to develop the secondary progressive course within a period of 10-25 years (Weinshenker et al., 1989; Noseworthy et al., 2000).

Progressive-relapsing and relapsing-progressive MS subtypes are also discussed in the literature (Kremenchutzky et al., 1999; Tullman, et al., 2004). These subtypes represent a combined course of primary-progression with some relapses (Kremenchutzky et al., 1999 suggest this be considered a primary progressive course), and a relapsing course with some progression, (considered a secondary progressive course; Kremenchutzky et al., 1999). Both the progressive-relapsing and relapsing-progressive descriptors attempt to capture the blurring between progressive and relapsing courses. There is some disagreement about their utility. Kremenchutzky et al., (1999) suggest they not be used, but at the same time recognise that 28% of people with a primary-progressive diagnosis do have relapses. Others suggest that the progressive-relapsing MS subtype is important because it has implications for drug treatment; a relapsing component to the disease course often being a requirement for the prescription of disease modifying drugs (Tullman, et al., 2004).

Finally, benign MS is a classification often given to those people who have little progression, or little or no acquired disability over a long period after diagnosis, though there is variability in the definition (Noseworthy et al., 2000; Pittock & Rodriguez, 2008). Follow-up studies over 10 and 20 year periods suggest that low disability, at over 10 years after diagnosis, is a good predictor of future disability, with up to 72% of benign MS patients remaining low in disability in two studies (Sayao, Devonshire & Tremlett, 2007; Pittock & Rodriguez, 2008). However longer-term follow up, has suggested that only 15% may actually remain benign (Costelloe et al., 2008).

MS affects approximately twice as many women as men (Tomassini & Pozzilli, 2009), an unexplained phenomenon noted in other apparently autoimmune diseases (Compston & Coles, 2002). Recently there has been some suggestion that the incidence is increasing in women (Orton et al., 2006). There are also some gender differences in MS subtype and disability prognosis, with males tending to have poorer outcomes (Tomassini & Pozzilli, 2009). Disease diagnosis peaks during the 3rd decade of life (Compston & Coles, 2002), with onset typically occurring between 2^{nd} and 5^{th} decade (Noseworthy, et al., 2000; Jacobs et al, 2000; Haussleiter, Brune & Juckel, 2009). While generally affecting adults, MS has also been diagnosed in children as young as 10-months old, though it is considered extremely rare before the age of 10 years; generally accepted onset is considered possible from 14-45years (Eraksoy, 1999), with between two and five percent of people experiencing their first clinical symptom before 16 years of age (Ness et al., 2007)

1.2.1. Incidence and Prevalence of Multiple Sclerosis

Multiple Sclerosis affects about 1 in 1000 people in the western world (Sadovnick & Ebers, 1993), with wide variations in prevalence relating to geographical location (Ebers, 2008). Increases in prevalence in the UK are documented in a number of studies, ascribed in part to better case ascertainment and differing diagnostic criteria for a diagnosis of MS (probable, clinically definite, single clinically isolated syndromes etc.; Robertson et al., 1996; Richards et al., 2002; Compston et al., 2006). The National Collaborating Centre for Chronic Conditions (2004) suggests an annual incidence of 3.5-6.6 new diagnoses, per 100,000 in England and Wales, and a prevalence of between 100-120 people per 100,000 (Richards et al., 2002; National Collaborating Centre for Chronic Conditions 2004). This equates to between 1,820 and 3,380 new diagnoses per year, and a population total of between 52,000 and 62,000 people with MS in England & Wales. Scotland has a higher prevalence than the rest of the UK (Shepherd & Summers, 1996), giving, in the UK as whole, estimations for the population of people with MS of about 85,000 (MS Trust, 2008). Many UK studies’ findings regarding prevalence fit with these figures (Shepherd & Summers, 1996; Robertson et al., 1996; Ford et al., 1998). Estimates for the year 2000 provided by Richards et al., (2002), suggested an incidence of 3.5-3.8 per 100,000 for England and Wales, and a prevalence of 92-104, per 100,000 people, for the south of England.

One prospective study, recording all incident cases of neurological disorders in 13 GP practices in London, UK, generated a sample of 100,230 people who were followed for the development of a neurological disorder over an 18-month period (MacDonald et al., 2000). The authors found an incidence of MS (calculated as number, per 100,000, per year) of 2 (95% confidence interval was 2-3).

1.2.2. The Economic Impact of Multiple Sclerosis.

MS is the leading cause of chronic neurological disability among working-age people in developed countries (Haensch & Jörg, 2006). As a chronic illness, the costs of MS evolve over time, and secondary disability may increase health care consumption and reduce economic output (Richards et al., 2002). Total cost of MS for the UK, has been put at about £1.3- £1.5billion for 1994 (Holmes, Madgwick & Bates, 1995), three years prior to the licensing of disease modifying treatments, which have been proposed to cost £8,000 per person, per year (Department of Health, 2002).

A recent cost and quality of life study for the UK (Kobelt et al., 2006), with 80% of respondents between 40 and 69 years, found that the average age at diagnosis was 38.8 years (standard deviation (SD) = 10 years) and mean age at first symptom was 32.2 years (SD = 10.4years); 28% were employed. Total mean cost per participant was estimated at £30,263 per year, with the cost of a relapse estimated at £1,164. These costings included medical, non-medical and societal costs directly attributable to MS.

Costs may thus increase with disease severity, but some have proposed that more rehabilitative models of care have implications for potentially reducing costs and for increasing patient satisfaction (Rotstein et al., 2006). Such rehabilitative models of care are inferred in the National Service Framework guidelines, which aim to guide the development of improved self-management, rehabilitation, equity of access and coordinated service provision for people with chronic neurological disease (Department of Health, 2005).

1.2.3. Causes of Multiple Sclerosis

Despite increasing understanding of the mechanisms of damage caused by MS, the cause, or causes, of MS are unknown, with current theories proposing the interplay of genes and environment (Compston & Coles, 2002). Specific environmental agents may include exposure to common viral agents, such as the Epstein-Barr virus (Ascherio & Munger 2007a). Current investigations in part focus on the contribution of genetic susceptibilities, evidenced by the clustering of MS within families (Giovannoni & Ebers, 2007). Additionally, there is an assumption that such susceptibility becomes meaningful only in the context of environmental triggers, and the age of exposure to as yet unknown environmental factors (Poskanzer et al., 1976; Gale & Martyn, 1995). Geographical patterns in the prevalence of MS could be considered supportive of genes, environment, or both, as causative agents, though migration evidence suggests that risk declines when people migrate away from high prevalence areas, which might support an environmental trigger (Gale & Martyn, 1995; Ascherio & Munger, 2007b). Noseworthy et al., (2000:942) in an attempt to summarise the likely multi-factorial nature of MS, characterises it as *'an immune-mediated disorder that occurs in genetically susceptible individuals'*, with susceptibility relating to some environmental trigger.

1.2.4. Diagnosis

The main framework for diagnosis of MS is by means of clinical guidelines (e.g. McDonald criteria; McDonald et al., 2001) which propose that in order to diagnose *Multiple* Sclerosis, white matter disease should be disseminated in time (lesions appearing between 30 days to 3 months apart) and space (in different parts of the brain and spinal cord; McDonald et al, 2001; Poser & Brinar, 2001; Polman et al, 2005; Wingerchuk & Weinshenker, 2008). Many authors however suggest that taking a history, and making a good clinical examination, is sufficient to make a diagnosis in many instances (National Collaborating Centre for Chronic Conditions, 2004; Polman et al., 2005).

In general, the guidelines for diagnosis are permutations of clinical signs (McDonald et al, 2001), history given by patients (Poser & Brinar, 2001), MRI findings (brain and spinal cord) and evidence of the by-products of the assumed inflammatory demyelinating process in the cerebrospinal fluid (McDonald et al, 2001; Polman et al, 2005). In comparison with previous guidelines for diagnosis, the McDonald criteria reflect technical advances, and give an increased importance to MRI findings (Poser & Brinar, 2001). One of the proposed benefits of the increasing use of MRI to assist with diagnosis is to reduce the delay in providing a diagnosis to patients (National Collaborating Centre for Chronic Conditions, 2004). This may be because it is effective at finding disseminated lesions (either in time or space), reflecting the most common mode of disease onset, relapsing-remitting MS.

The guidelines also take account of differing clinical presentations, including the typical relapsing-remitting presentation, where time and space dissemination is key, and primary-progressive presentation, where a one-year history of progression and relevant MRI findings is warranted. For the so-called monosymptomatic disease (alternatively termed clinically isolated syndrome), where dissemination in time and space may not, yet, be evident, both MRI and CSF findings are relevant. A ‘wait and see’ approach is often taken to see if a relapse occurs, potentially explaining the often-reported long wait for a diagnosis. Follow up MRI scans are recommended in such cases to monitor for dissemination (Polman et al., 2005; McHugh, Galvin & Murphy, 2008). The guidelines are also specific that a relapse should include neurological symptoms likely to be caused by demyelination, that it lasts over 24 hours, and that subjective report is corroborated by clinical signs (McDonald et al, 2001; Polman et al, 2005).

1.3. Symptoms and effects

Given that the demyelination can occur throughout the central nervous system, the symptoms of MS can be varied. There seem to be some proclivities during the course of the neurodegeneration however, with some symptoms being more common than others, and more common at different stages of the illness (Paty, 2000; Noseworthy et al., 2000; El-Moslimany & Lublin, 2008). An outline of the main categories of symptoms is presented now, with greater attention given to mood and cognitive symptoms, because these will be the focus of the study.

1.3.1. Sensory dysfunction

The most common symptoms of Multiple Sclerosis are disorders of sensory function (Paty, 2000), which can include changes in the sensory experience, even with intact motor function (Paty, 2000). This might present as a general clumsiness or complaints of dropping things. Other sensory symptoms related to optic neuritis (inflammation and demyelination of optic tract) are blurring of vision, blindness, typically in the central field of vision, visual distortions, and changes in contrast sensitivity or colour vision (Ashworth, Aspinall & Mitchell, 1989). Visual disturbances are common in MS, affecting over 50% of people (Anderson & Cox, 1997) and sub-clinically potentially more (Lycke, Tollesson & Frisén, 2001).

Pain, or neuropathic pain, as a primary symptom, can be experienced as acute paroxysmal (transient, intense, pain symptoms), as burning sensations, or neuralgia along the trigeminal nerve distribution (Polman et al., 2006; El-Moslimany & Lublin, 2008). Chronic radiating pain along a specific nerve distribution (so called radicular pain) is not uncommon (Ramirez-Lassepas et al., 1992). In addition to the experience of disordered sensory experiences, sensory dysfunction may also impact on bladder and bowel control, where reduced sensory appreciation contributes to problems with continence.

In general, sensory symptoms tend to be rapid in onset (Paty, 2000) and initially (early in a relapsing-remitting course), may resolve over a few days, or with a course of steroid treatment (Noseworthy et al., 2000). As the disease progresses, or as more relapses occur, some sensory symptoms may become permanent.

1.3.2. Motor Dysfunction

Disorders of motor function are also common; tremor, ataxia (incoordination), dysarthria (motor articulation problems), spasticity and weakness in muscles are frequently experienced (Calabresi, 2004). These motor disorders are not limited to the limbs, trunk or orofacial muscles and so can, for example, affect bladder function; increased bladder muscle tone can lead to emptying difficulties. For some, notably with more severe ataxic (cerebellar) disorders, motor control rather than muscle weakness or stiffness, can cause profound limitations, making head, limb, eye and trunk control so difficult that most aspects of activities of daily living are impaired (DeSouza & Bates 2004).

Motor disorders of the eyes are also frequently encountered, typically, though not solely, relating to brainstem (cranial nerve) or visual tract demyelination. Eye muscle control - e.g. inter-nuclear opthalmoplegia, leading to conjugate gaze disorders, oculomotor palsies leading to diplopia (double vision), and poor control of saccadic eye movements are all seen (Noseworthy et al., 2000). Unlike sensory symptoms, motor disorders tend to evolve more slowly (Paty, 2000).

1.3.3. Fatigue

Fatigue is a third, and very common symptom of MS, experienced by over half of people with the disease (Bakshi et al., 2000; Krupp & Christodoulou, 2001). It is both one of the most commonly reported and top-rated worst symptoms of MS (Multiple Sclerosis

Society, 1997; Brañas et al., 2000). While the experience of fatigue may be contributed to by poor sleeping, depression, or as a side effect of medication, it is primarily seen as a symptom in itself. MS-related fatigue is a chronic ongoing problem, experienced daily in most people with MS (Freal, Kraft & Coryell, 1984; Brañas et al., 2000; Rammohan et al., 2002).

The primary experiences include the abrupt and severe 'MS fatigue' called *lassitude* (DeLuca, 2006; Schapiro, 2007). In addition, local fatigue, affecting specific muscle groups is common (Brañas et al., 2000; Schapiro, 2007). It is therefore considered qualitatively different from 'normal' fatigue and leads to a significant, frequently sudden, curtailment of function (Brañas et al., 2000). The causes of fatigue in MS are not fully understood, though there are likely to relate to structural (demyelination) and biochemical changes because of MS (Bakshi, 2003; Schapiro 2007).

1.3.3. Autonomic Dysfunction

Another cluster of symptoms relates to autonomic dysfunction, and includes aspects of bowel (gastric emptying) and bladder function (mainly detrusor muscle control). Sexual dysfunction (e.g. impotence, erectile and ejaculatory dysfunction in men, anorgasmia and reduced lubrication in women), as well as low blood pressure (orthostatic intolerance) are common results of autonomic dysfunction (Haensch & Jörg, 2006).

1.4. Cognitive Dysfunction in Multiple Sclerosis

Cognitive impairment is noted in up to 65%, of people with Multiple Sclerosis (Rao, et al., 1991; Rao, 1995; Deloire et al., 2006), the proportion being in part dependent on the sample and MS subtype (Rao et al., 1991; Denney, Sworowski & Lynch, 2005). Progressive forms of MS display a greater proportion of cognitive impairment compared to relapsing-remitting course (Beatty, et al., 1989) meaning that samples, and how they are acquired, can differ in prevalence of impairment.

In MS, there is a proclivity for white matter changes to be periventricular, within the deep white matter of the frontal lobes and around the corpus callosum. As a result, some similarities in cognitive profile have been noted among people with the disease (Brownell & Hughes, 1962; Hannay, et al., 2004). Broadly, the most common pattern of cognitive deficit is seen in two or three domains - attention/processing speed, memory and executive functions (Fischer, 2001; Benedict et al., 2002; Hannay et al., 2004; Nocentini et al., 2006).

In respect of MS subtype, one study of patients with relapsing-remitting MS suggested about 39% had impairment in information processing and memory domains (Nocentini et al, 2006). For primary- and secondary-progressive MS, there is more severity of impairments as well as a higher incidence of impairment (Beatty et al., 1989; De Sonneville et al., 2002). One reason for this is likely to relate to the relapsing-remitting samples being younger, and perhaps more importantly, having less accrued disability or lesions, or having MS for a shorter time. Not all studies find relationships between cognitive abilities and disability; this may be, in part, because the main assessment of disability (the Expanded Disability Status Scale) is biased towards physical disability (Hoogervorst et al., 2001). Others suggest that there are relationships only between specific aspects of cognition (e.g. information processing abilities) and disease duration and disability (De Sonneville et al., 2002).

Cognitive impairment has been likened to the presentation of subcortical dementia (Brassington & Marsh, 1998) where cognitive slowing, forgetfulness, reduced insight and depression are typical features. A different view, that of *disconnection* within brain structures, resulting in generalised rather than specific impairment is also proposed (Deloire et al., 2006). Both views implicate damage to the brain-wide subcortical networks that support attention and information processing (Brassington & Marsh, 1998). It is considered uncommon for so-called ‘cortical signs’ to appear (e.g. aphasia, apraxia, agnosia; Lezak, Howieson & Loring, 2004:244). Given the focus of the proposed study, outlined in the introduction, performance characteristics of samples of people with MS in the areas of information processing, memory and executive function are considered.

1.4.1. Information Processing

In the unimpaired human brain, despite massive capacity for parallel processing, there are major limitations on both in terms of speed (refractory periods between one process and its repetition) and capacity, or the number of parallel computations (Marois & Ivanoff, 2005). Reflecting this, information processing in MS has been investigated from the point of view of speeded performance (Gontkovsky & Beatty, 2006) and capacity (Archibald & Fisk, 2000; McCarthy et al., 2005).

Relatedly, Chiaravalloti et al., (2003) and Archibald et al., (2004) question what the construct of information processing actually describes, not just in MS. Archibald et al., (2004:1562) consider information processing capacity as '*cognitive speed, complex attention and working memory*'. In a study in people with chronic fatigue syndrome, Chiaravalloti et al., (2003) proposed that three elements, which they termed *simple speed and reaction time*, *complex information processing speed* and *working memory ability* could be separated. In MS samples too, information processing remains difficult to characterise, because of the relationships between it and working memory and executive abilities (Drew, Starkey & Isler, 2009).

Nonetheless, information processing is a key and pervasive deficit in MS, regardless of disease subtype, or whether it is measured in intentional or automatic processing tasks (DeLuca et al., 2004; Denney, Sworowski & Lynch, 2005; Olivares et al., 2005; Sepulcre et al., 2006; Parmenter et al., 2007). A range of task manipulations to investigate both automatic and controlled processing abilities does however suggest that task complexity is positively associated with increased difficulties with more complex aspects of information processing (De Sonneville et al., 2002).

Changes in information processing with normal ageing have been proposed to mediate performance on a range of more complex cognitive operations (e.g. Stokx & Gaillard, 1986; Salthouse, 1996; Bunce & Macready, 2005). Conversely, reduced frontal or executive functions have also been offered as an explanation for cognitive ageing (Bugaiska et al., 2007). Both bottom-up and top down mechanisms are therefore proposed to explain cognitive limitations in ageing studies, with information processing abilities considered a bottom up influence. In MS, the focus has typically been on the widespread deficits in information processing, notable speed, as a key limiting factor in a range of cognitive operations (DeLuca et al., 2004).

The true impact of these deficits might therefore be seen as performance decrements in attention-demanding, speeded, controlled cognitive processing (Schneider & Schiffrin 1977; Grafman et al., 1991; Salthouse, 1996). Management strategies focusing on reducing time pressure maintains performance accuracy in MS, supporting the proposal that information processing acts as a mediator of performance in a range of cognitively demanding tasks (Salthouse, 1996; Demaree et al., 1999; MacNiven, et al., 2008). The implications in more complex cognitive tasks are typically seen in a new learning decrement and impaired working memory performance (Archibald & Fisk, 2000; Chiaravalloti et al., 2003; Lengenfelder et al., 2006; Drew, Starkey & Isler, 2009). Many have suggested that speed of processing specifically is the main feature of this common deficit (Demaree et al., 1999; DeLuca et al., 2004; Goverover et al., 2007). Despite this, however, O'Brien et al., (2008:766) suggest that 'a significant omission' in the MS literature is its consideration.

In summary, information-processing abilities are frequently impaired in MS and there appears to be an upward impact on other cognitive abilities. Performance across a range of processing tasks may be factorially separable into complex speeded information processing and working memory deficits, but which of these has primacy in mediating performance on learning tasks remains debated. There are limitations in the literature in terms of a shared understanding of the relationships between information processing capacity, speed, attention, working memory and new learning. While information processing may impact on memory performance in MS, especially in terms of acquisition processes, memory abilities themselves have been demonstrated to be impaired in up to 60% of people with MS (Rao et al., 1993; Brassington & Marsh, 1998).

1.4.2. Memory

As Brassington & Marsh (1998) discuss, memory is not a unitary function, so patterns of performance, or impairments in components of memory need to be considered. Memory may be impaired in respect of some of the processes associated with its functioning - acquisition, retention or retrieval as well as in terms of memory type - working memory, episodic or semantic memory (Cermak, 2000; Tulving & Craik, 2000)

People with MS appear to have difficulties with the acquisition stage of new learning, so that while incremental learning takes place with repeated exposure to material, the learning curve is reduced in terms of amounts acquired at each stage (Rao et al., 1989; Diamond et al., 1997). Some have considered this a feature of reduced information processing capacity (Archibald & Fisk, 2000; Chiaravalloti et al., 2003; DeLuca et al., 2004), but the failure could also relate to poor encoding, ineffective strategy or organisation, unrelated to the speed of presentation (Arnett et al., 1997).

This latter acquisition impairment has been associated with contributions of the prefrontal cortex (Petrides, 2000), and executive abilities (Arnett et al., 1997; Canellopoulou & Richardson, 1998). Lezak, Howieson & Loring (2004:251) discuss this aspect of memory performance in MS; the ‘frontal’ nature of impairments meaning, people with MS tend to less often use strategic approaches to encode information. They use less semantic clustering as a strategy for organising to-be-remembered information (Raymond et al., 1987; Arnett et al., 1997) or have difficulty with strategy use (Carroll, Gates & Roldan, 1984). This is notable if the organisation into semantic categories is not obvious (Arnett et al., 1997), perhaps suggesting failures to consider the to be learned information.

These problems with processes optimising memory performance, rather than doing the remembering itself, are considered to be executive, or at least frontal contributions to memory performance (Petrides, 2000). Frontal contributions might also be helpful in preventing spurious associations being learned in memory contexts (Gisiger, Kerszberg & Changeux, 2005), highlighting the important role in monitoring at both encoding and retrieval. At a more general level, active monitoring, in respect to goal state, is a consideration too (Van Overschelde, 2008). This perspective on memory performance might be considered to reflect the top-down directing of memory-oriented behaviour by executive processes.

An alternative, or perhaps complimentary, explanation might be that this encoding process constitutes effortful or controlled, intentional processing. This could be constrained by the reduced information processing capacities common in MS. A number of permutations of acquisition failure could therefore exist; a processing speed deficit, ineffective strategy use, or failure in strategy use, because of limited processing resource. The debate about top-down and bottom-up contributions to cognitive performance in the ageing literature may have something to offer, with caveats, in interpreting results in memory acquisition failures for this group (e.g. Bunce & Macready, 2005; Salthouse, 1996; 1991). Of particular interest in this study is whether similar conceptual findings - processing speed accounting for variance in memory performance - might also apply to metamemory.

Retrieval is a second process relating to memory performance and can be impaired because information is forgotten, or because of impairments in retrieval processing (Laming, 2009). A key question in the MS literature is whether people with MS tend to forget more than those without, once information has been learned (that is, by controlling for acquisition deficits; Diamond et al., 1997).

In MS, while forgetting (measured by quantity retrieved) does take place at a higher rate than in control samples (Calabrese, 2006); acquisition deficits may be the more significant deficit (Diamond et al., 1997; Olivares et al., 2005). One difficulty with interpreting retrieval failures is whether a retrieval failure indicates true forgetting or, instead, failures at retrieving target items (Laming, 2009). On tests of recognition memory, which attempts to separate *forgotten* from *failed to retrieve*, people with MS can perform at a normal level (Rao et al., 1991), suggesting failure at retrieval, rather than forgetting, as the difficulty (Grafman, Rao & Litvan 1990; Calabrese, 2006). The implications of failure to retrieve, as opposed to forgetting, probably relate to the processes surrounding retrieval. These again implicate frontal functions, and include poor organisation at retrieval (Stuss et al., 1994; Petrides, 2000) or ineffective retrieval strategies, leading to self-interfering effects (Butters et al., 1986; Laming, 2009).

In summarising memory abilities in MS, it is necessary to consider the type of memory that is typically assessed. Most tests of memory require new learning trials, meaning that episodic memory has a greater research base (Beatty et al., 1988; Paul et al., 1997). Evidence has been discussed which suggests that acquisition failure does occur, but when mitigated, forgetting is less severe than might be perceived. Additionally, other processes supporting both acquisition and retrieval, implicating executive functions, may be impaired. As executive function can be impaired in MS, and given the focus of this study on memory monitoring, attention to performance in samples of people with MS is appropriate.

1.4.3. Executive Function

Executive functions are implicated in performance of tasks that are non-routine, novel, complex, or for which typically used schemas are unsuitable to generate appropriate behavioural routines (Norman & Shallice, 1986; Shallice & Burgess, 1991; 1993; Shallice, Burgess & Robertson, 1996; Godefroy, 2003). Disorders of executive function are associated with both frontal lobe damage, and disconnection (Burgess, 2000; Goldberg, 2001). Executive function is an umbrella term for a range of cognitive operations serving ongoing goal-directed behaviours (Rabbitt, 1993), including generation of possible actions, inhibition of others, selection and adaptation of plans. These goal-directed behaviours could include action, speech or reasoning (Fuster, 2002). While this definition might suggest that complex tasks are, by their nature, 'executive', Rabbitt (1997) cautions against this interpretation:

'The key distinction seems to be between situations in which the person must, for the first time, recognise, evaluate, and choose among a variety of alternative options and those in which a single effective behaviour sequence, which has been previously identified, and instantiated by practice, is run off without the need to propose and evaluate alternatives' (Rabbitt, 1997:3)

While the function of memory might be to make us independent of our immediate stimulus environment (Watkins, 1990), the function of executive abilities is to generate, initiate and control behaviours (real or imagined), independent of the external environment, that is, to anticipate (Goldman-Rakic, 1994; Okuda et al., 2003). In not being dependent on the external environment, there are options to adapt behaviours if the environment does not behave as expected. This flexibility requires the ability to adequately represent or characterise a problem, terminate or inhibit ongoing behaviours, adapt them or generate new plans in an attempt to control future environment or task possibilities, and be strategic about future behaviour (Shallice, Burgess & Robertson, 1996; Rabbitt, 1997). One approach to thinking about the mechanics of executive function is that of the Supervisory Attention System (SAS) proposed by Shallice (1988). The SAS is a system that is proposed to manage the selection of various action schemas to guide goal-directed behaviour through a process of monitoring and modulation of the cognitive system (Shallice, 1988).

The range of behaviours associated with executive function include preparation (preparatory set, or planning; Fuster, 2002); inhibition (e.g. of internal drives, of inappropriate behavioural schemas, or reactions to task-irrelevant external stimuli; Fuster, 2002; Godefroy, 2003); a problem detection, characterisation and solving ability (Shallice 1988) and flexibility and strategic thinking to adapt to changes (Burgess, 1997). One view is that measures of fluid intelligence best capture executive performance (Duncan, Burgess & Emslie, 1995; Zook et al., 2004).

Performance of people with MS on tests of executive function suffers from the same limitations from which all tests of executive function suffer - limited ecological validity and task-impurity (Burgess, 1997). Measures of executive function typically correlate with other cognitive indices, such as memory (as in the Wisconsin Card Sorting Test; Burgess, 1997) or visuospatial functioning (as in the Trail Making Test; Burgess, 1997:88). A third limitation relates to the potential for fractionation or dissociation of deficits in executive function (Miyake et al., 2000; Duncan et al., 1997), meaning that samples may perform differently on different tasks, because different dimensions are being tested.

Executive dysfunction is commonly reported in people with MS (Arnett et al., 1997; Benedict et al, 2002; Clemmons et al., 2004; Wachowius et al., 2005). A range of different measures of executive function have been used, including rule-based card sorting tests (Beatty et al., 1989; Beatty & Monson, 1991; Rao et al., 1991; Benedict et al., 2001), which suggest that generating novel strategies (and dropping ineffective ones) might be a key executive skill (Lezak, Howieson & Loring, 2004; Calabrese, 2006).

Other tasks proposed to load on planning aspects of executive function, include tower solution tasks (Tower of London, Tower of Hanoi; Zook et al., 2004); blocks or discs have to be moved across a set of pegs following various rules to create a specified pattern. Performance of samples of people with MS on such tasks has been investigated in a number of studies (Randolph, Arnett & Freske, 2004; Denney, Sworowski & Lynch, 2005). These suggest that planning time (from presentation to first move) and execution time (time from first to last move) can be impaired, especially the latter. Of interest, the score (number of moves) for the tower task in Denney, Sworowski & Lynch, (2005) was similar for MS participants and controls, once groups were equated for age, gender, education, depression and fatigue. As was suggested in a previous study (Denney et al., 2004), speeded performance was the limiting issue. This is of interest because it lends some support to the suggestion that there may be an underlying mediation of processing abilities, notably speed, on performance in this domain, if time is used as a measure (Demaree et al., 1999; Archibald & Fisk, 2000; Denney, Sworowski & Lynch, 2005). Using time to measure executive abilities is likely to be necessary to be sensitive to ability, but where processing-time and executive-time separate may be a challenge to discern. This revisits the possibility of information processing abilities serving a mediating role in such executive tasks, similar to that proposed in memory performance.

One cognitive skill, which relates strongly to both information processing and executive ability, is working memory (Baddeley, Hitch & Bower 1974; Baddeley, Della Sala & Robbins 1996). The inclusion of working memory tasks as indices of executive function is sometimes considered an issue of theoretical orientation (Lezak, Howieson & Loring, 2004), possibly reflecting the different traditions in how working memory is researched in the United States (top-down, executive control focus) versus Europe (more bottom-up; Baddeley & Hitch, 2007:7). It can complicate the issue of what is meant by executive function (Zook et al., 2004). One example is the proposal that working memory, mental shifting and inhibition might be three key indicators of executive abilities (Pennington et al., 1996; Miyake et al., 2000).

Statistical approaches may be one way of separating these conceptual domains. The use of confirmatory factor analyses or hierarchical regression models may be able to indicate where some dimensionality might exist in individual samples (Miyake et al., 2000; Zook et al., 2004). A benefit of this approach is that a measure's construct validity is assessed based on the sample under study, and not from the typically unimpaired derivation sample. The difficulties with separating the concepts of executive function and working memory are reflected in the literature not just in terms of selection of measures. Neuroimaging evidence suggests shared anatomical substrates for working memory and executive functions (Bayliss et al., 2003). For this reason it is germane to consider how working memory relates to executive function.

The concept of working memory developed most notably in the work of Baddeley & Hitch (Baddeley & Hitch, 1974; Baddeley 1996; Andrade, 2001a) and is presented as having both executive and non-executive elements (Baddeley & Hitch, 2001, 2007). It is proposed to include two processes - the ability to maintain information in mind (termed 'on-line') long enough to use it and the ability to manipulate that information so as to optimise it for use. Optimising might involve reorganisation based on features of the stimuli, such as semantic categorisation to better remember it (Andrade, 2001a; Daneman & Hannon, 2007). The Baddeley & Hitch working memory model has an organising system, the central executive, which manages the operation of two 'slave' systems, each of which manages different modalities of stimuli (Baddeley & Hitch, 2007:14). The visuo-spatial sketchpad manages visual and spatial to-be-maintained information (e.g. in the use of imagery mnemonics) and the phonological loop manages auditory and language stimuli.

The functioning of working memory therefore requires ongoing attention or active maintenance (Miyake & Shah, 1999). The resources available to the active maintenance are limited and if they are not applied, or if they are interrupted by interference, the system can fail (Richardson, 1996).

More recently, a third slave component was proposed in which a buffer mechanism operates with episodic memory to allow for integration of the 'on-line' store and long-term memory (Baddeley, 2000). The main proposed function of this buffer is as a workspace for the integration of incoming and stored memory to update, adjust or expand experience and for the use of stored knowledge to inform the organisation of information in working memory, e.g. for organising material semantically or into 'higher order representations' (Baddeley & Hitch, 2007:7).

A co-ordinating role in the operation of the slave systems (phonological loop, visuospatial sketchpad and episodic buffer), is played by the central executive component of working memory, for example in updating the contents of working memory on a continuous basis (Lowe & Rabbitt, 1997; Brownell & Friedman, 2001). The links between working memory and executive function are based on strong associations between anatomical substrate in functional imaging studies (Bayliss et al., 2003; Oberauer et al., 2007) and in terms of the proposed function of the central executive component of working memory being to 'simultaneously maintain and process goal relevant information' (Conway et al., 2007:3).

Limited capacity is a feature of working memory, and this and its susceptibility to the effects of interference, has meant that as well as being considered an appropriate index of executive abilities, it has also been considered a route to indexing attention and processing abilities (Beatty, 1998; Fisk & Archibald, 2001; Baddeley & Hitch, 2007). As mentioned, the differentiation in part is based on the theoretical orientation of researchers. For example, in reviewing comprehensive cognitive assessment in MS, the Peyser et al., (1990) guidelines for a core battery for neuropsychological assessment suggest a domain of 'information processing and working memory' separate to an assessment of 'reasoning and executive functions'. In developing their own minimal assessment guidelines Benedict and the consensus group on neuropsychological assessment in MS (Benedict et al., 2002) agreed that processing speed and working memory was an important sphere for assessment, separate from executive function. However in studies in the MS literature, what are typically considered indices of working memory are used as surrogates for executive function (e.g. Letter Number Sequencing task in Randolph, Arnett & Freske, 2004).

Archibald & Fisk (2000) suggest that while working memory impairments might be seen in MS, unlike information processing, they are not a pervasive deficit. Their findings suggest that impairments in working memory may indicate a person having reached a threshold level of information processing deficit, as might be associated with greater levels of cognitive dysfunction, typical of secondary-progressive as opposed to relapsing-remitting stage of the disease.

The domains of memory, information processing and executive function have been found to be frequently impaired in MS samples and an intuitive association with compromised learning has been confirmed experimentally in MS (Demaree et al., 1999, Arnett 1999; Chiaravalloti et al., 2005; Goverover, Chiaravalotti & DeLuca, 2008). However, only to a limited extent have these cognitive factors been investigated in respect of the monitoring of learning (Beatty & Monson, 1991). In addition, relevant non-cognitive factors might also impact on appraisals of ability (Conklin, Strunk & Fazio, 2009). As mood disorder is also common feature of MS, and may also be relevant to accurate memory appraisal (Randolph, Arnett & Freske, 2004; Julian, Merluzzi & Mohr, 2007), it is explored here.

1.5. Affective Disorders in MS

Mood disturbances in MS were noted by Charcot in 1879 (Talley, 2005) and it remains difficult to ascertain if they relate to psychosocial, reactive, iatrogenic or organic causes (Minden & Schiffer, 1993:40; Mohr & Cox, 2001; Surguladze, Keedwell & Phillips, 2003; Goldman Consensus Group, 2005; Gold & Irwin, 2006; El-Moslimany & Lublin, 2008). Two key affective disturbances are discussed in the literature, depression and euphoria.

1.5.1. Depression

Depression in MS is quite likely to contain both organic and psychosocial components (Ford & Naismith, 2006). An additional causative component may be iatrogenic; one listed potential side effect of some disease-modifying (e.g. interferon) treatments being depression (Ford and Naismith, 2006). Steroid treatments have also been associated with depression, as well as hypomania. One difficulty is dissociating the concurrent impact of a relapse, for which they might be prescribed, on mood (LaRocca, 2000; Mohr & Cox, 2001; Patten & Metz, 2001, 2002). Studies seem to confirm incidence of depression is associated more with cerebral, compared to spinal, demyelination (Ford & Naismith, 2006), perhaps supporting an organic rather than reactive contribution.

One caveat in associating disability type and depression in this way, is that studies may tend to recruit more acutely disabled (i.e. relapsing), than chronically disabled, participants (Minden & Schiffer, 1993). Other studies have shown that longer-standing illness might be associated with more successful adjustment, appearing to conflict with the suggestion that disability and depression might be correlated (Brooks and Matson, 1982). Overall, a direct relationship has not been consistently established between depression and type of disability e.g. motor versus sensory (Minden, Orav & Reich, 1987), type of MS (Minden, 2000), duration of symptoms, or cognitive function (Minden & Schiffer, 1990, 1991, 1993; Patten & Metz, 1997). More recent studies have suggested that depression is negatively related to cognitive impairment (Landro, Sletvold & Celius, 2000; Haase et al., 2004).

Lifetime prevalence of major depression in people with MS ranges from 37% to 54% (Minden & Schiffer, 1993; Mohr & Cox, 2001; Feinstein, 2007). To some extent these findings may have been biased by the sample being current healthcare users (McGuigan & Hutchinson, 2006; Haussleiter, Brune & Juckel, 2009), and by the tools used to measure depression (e.g. Beck Depression Inventory 1st edition). Previous versions of the Beck Depression Inventory have been suggested to confound mood with other neurological symptoms, notably fatigue, in MS (Nyenhuis et al., 1995; Goldman Consensus Group, 2005; Ford & Naismith, 2006: 263).

In a large MS sample, Chwastiak et al., (2002) suggest that ‘clinically significant’ depression was present in 42% of respondents, high in comparison to other studies of prevalence in the general population (3-9%) or primary care patients (10-15%), and, some suggest, even other neurological samples (Schubert & Foliart, 1993). Point prevalence for major depression for MS-clinic samples was reported as 14% according to one review panel (Goldman Consensus Group, 2005). Yet one study, from outside North America, which attempted to control for sampling bias and using the revised Beck Depression Inventory II (BDI-II), reported a point prevalence of 28% for moderate or severe symptoms of depression (McGuigan & Hutchinson, 2006). The moderate to severe depression symptom group had a shorter duration of illness, but similar levels of disability on the Kurtzke Expanded Disability Status Scale (EDSS: Kurtzke, 1983) to the non-depressed or mild depressed members of the sample. This might offer some support for an adjustment-related process in determining current depression.

Another study proposed that severity of illness, and not length of time of illness, was associated with depression (Chwastiak et al., 2002). Patten et al. (2003), in a population-based sample, suggested an increased period prevalence (1 year) for people with MS, compared to those without MS or with other long-term conditions. In a number of studies, being younger (Chwastiak et al., 2002) and in some studies, being female (Ford & Naismith 2006; McGuigan & Hutchinson, 2006) was also associated with higher prevalence. In the UK, a postal survey of people with MS (Sollom & Kneebone, 2007), employing a widely used self-report tool for the assessment of depression (Centre of Epidemiologic Studies Depression Scale, or CES-D) found moderate to severe depression (scores >16) in 60% of respondents. Minden's (2000) suggestion that reliable predictors of depression in MS have not been identified appears generally true.

In terms of measurement of depression, the tools and the samples used are limiting factors when attempting to generalise to the MS population as a whole (Haussleiter, Brune & Juckel, 2009). One point, consistently made in the literature, is the difference between people with MS endorsing high numbers of depressive symptoms, such as fatigue, with measures such as the BDI and CES-D, and an actual diagnosis of a depressive disorder, based on full psychiatric interview and assessment (McGuigan & Hutchinson, 2006; Sollom & Kneebone, 2007). An additional debate is whether commonly used self-report measures reflect physical disability symptoms and so called neurovegetative symptoms of depression (Nyenhuis et al., 1995; Randolph et al., 2000; Goldman Consensus Group, 2005; Ford & Naismith, 2006:263).

Additional confounds are fatigue (Ford & Naismith, 2006:266), notably mental fatigue (Schreurs, de Ridder & Bensing, 2002), though the direction of the effect seems to be that mood state informs fatigue perception, at least based on the work of Mohr, Hart & Goldberg, (2003) who treated depression, leading to reductions in reported global fatigue.

The implication of high levels of depressive symptomatology in this group of people is that it may bias reasoning about abilities (Conklin, Strunk & Fazio, 2009; Strunk & Adler, 2009). More specifically, when asked to appraise aspects of their own ability or performance, some studies suggest that people, including those with MS, show a negative association between depressive symptoms and appraisals of cognitive ability (Randolph, Arnett & Freske, 2004; Julian Merluzzi & Mohr, 2007; Kit, Mateer & Graves, 2007). With more severe depression however, cognitive abilities may be impacted, as opposed to just underestimated (Diamond et al., 2008).

1.5.2. Euphoria

Another, apparently affective, disorder is MS-related euphoria or ‘euphoria sclerotica’ (Cottrell & Wilson, 1926:8). Historically, euphoria was considered a more common problem than depression, and understanding has shifted from it as a mood disorder to an understanding of it as frontal mediated disinhibition (Benedict et al., 2001). Minden & Schiffer, (1993) argue that the euphoria noted in some people with MS has a neurological basis. Unlike depression, is relates to disease severity and length, disability, a chronic progressive MS course, enlarged ventricles and more significant cognitive impairment (Surridge, 1969; Rabins, 1990; Minden & Schiffer, 1993; Fishman et al., 2004). The association between executive deficit and the often-described MS personality change has similarly been proposed to be an impairment of self-regulation (Benedict et al., 2001). The implied relationship between executive dysfunction and personality change is in the presentation of disorganisation, stubbornness, rigidity of behaviour and a tendency to poorly estimate performance or ability, that is to lack insight (Benedict et al., 2001; Lezak, Howieson & Loring, 2004).

1.6. Treatment of Multiple Sclerosis

All of the discussed symptoms associated to varying degrees with MS, present significant clinical challenges, aside from the daily living problems for the person themselves. Symptom management and treatment of MS falls into two broad categories: Medical management of symptoms, including modification of the disease process and second, attenuation of symptoms and disability, which includes neurological rehabilitation.

1.6.1. Medical Management

Broadly, pharmacologic management is offered for symptom control, secondary complications and toward the disease pathology itself. Symptom management discussed first, focuses on pharmacologic interventions for the range of symptoms presented earlier.

Probably the most commonly treated problems relate to disorders of a neuromuscular type - spasticity, stiffness, spasm, including paroxysmal spasm and associated pain (Calabresi, 2004). Thus, medications that act as skeletal muscle relaxants such as Dantrolene Sodium (Dantrolene), Tizanidine (Zanaflex), Baclofen (Lioresal) and Gabapentin (Neurontin), are often prescribed (Schapiro, 2006; Bhatia, 1999; Joint Formulary Committee, 2009). All of these medications have fatigue-related side effects in higher doses (Calabresi, 2004). Additionally, they may cause muscle weakness or have such a systemic effect on muscle activity as to reduce function (Young, 2000).

A second area of symptom management is in the management of fatigue. Within the context of depression, antidepressants may reduce the experience of fatigue (Schapiro, 2006). However the lassitude discussed earlier may respond better to stimulant-like medications such as Modafinil (Provigil; Rammohan et al., 2002) or with off-label use of Amantadine (Cohen & Fisher, 1989; Calabresi, 2004). Studies supporting the efficacy of pharmacologic treatment for fatigue have been inconclusive (Stankoff et al., 2005), possibly because there is little consensus about the measurement of fatigue or understanding of its mechanisms (Pucci et al., 2007).

While many medications are provided for the management of symptoms of multiple sclerosis, a number aim to directly influence the course of the disease. Generally they are derived, based on the presumed pathology of MS. Given that MS is, probably, an autoimmune inflammatory demyelinating disease, various treatments aim to attenuate the autoimmune and inflammatory component (e.g. steroids such as Prednisone, Prednisolone or Methylprednisolone). These medicines are typically used for acute relapses (Calabresi, 2004; Leary, Porter & Thompson, 2005). The other medications in this category are considered disease-modifying because they may have a long-term effect on disease progression, such as in reducing the numbers of relapses (Compston et al., 2006). There have been no studies of a truly long-term nature because the relative recency of their introduction (Association of British Neurologists, 2009)

One set of disease modifying drugs is the interferons, two classes of which (Interferon 1-a and Interferon 1-b) are used to 'induce the formation of neutralising antibodies' (Noseworthy et al., 2000:948). They suppress the autoimmune response thought to contribute to MS. A second disease modifying medication, Glatiramer Acetate (Copaxone), is proposed to help demyelination because it has a similar set of polypeptides to those comprising myelin itself (Calabresi, 2004), leading to what has been termed a 'bystander' suppression at the site of developing MS lesions (Chen et al., 2001). The proposed action is as a decoy to the inflammatory response against myelin (Miller et al., 1997). The effect of the drug appears to be in reducing the frequency of relapse (Noseworthy et al., 2000). The interferons can be prescribed for people with clinically isolated syndrome (Kappos et al., 2006), relapsing-remitting MS and secondary progressive MS. Glatiramer Acetate being recommended only for relapsing-remitting MS

It is in part because of these limitations on prescribing that MS disease subtype becomes an important issue. None of the above treatments is recommended for use in primary progressive MS, for which there is no proven disease modifying agent currently in use. However, anecdotal evidence suggests that there is some benefit to be gained from a medication called Naltrexone, in very low doses (Low Dose Naltrexone, or LDN; Gironi et al., 2008).

One ongoing issue with these medications is their cost-effectiveness; one year's treatment costs approximately £8,000 (Great Britain. Department of Health, 2002a). As a result of the National Institute of Clinical Excellence's review of cost effectiveness (Great Britain. Department of Health, 2002b), a risk-sharing scheme was set up with the pharmaceutical industry in 2002 by which the cost of interferons and Copaxone (Glatiramer Acetate) to the NHS may reduce in the future if they fail to show the projected impact on quality-adjusted life years (Bogglid et al., 2009). Under this scheme, only people meeting criteria developed by the Association of British Neurologists, are considered appropriate for treatment, and are followed to assess long term cost effectiveness of the drugs (National Collaborating Centre for Chronic Conditions, 2004; Association of British Neurologists, 2009). Side effects of the interferons typically include flu-like symptoms, irritation at the injection point, and can include exacerbation of mood disorder (Feinstein, O'Connor & Feinstein, 2002).

Pharmacologic management offers important avenues for the management of a range of MS symptoms, including neuromuscular problems, fatigue, continence, pain and acute relapses. Disease modifying therapies have been shown to be effective in reducing relapse rates. Because of their relative novelty (the first interferon for MS, Avonex, was licensed in the UK in 1997), benefits have not been established over truly long-term timescales, so their cost-effectiveness remains debated. Ultimately disease-modifying approaches are a tool to manage the disease, potentially reducing disability, or slowing the time course of disability accruement, rather than offering a cure.

1.6.2. Rehabilitation

Rehabilitation for people with progressive disease aims to optimise ability and potential, through symptom management and education (Wade, 1992; Wade & De Jong, 2000; Kesselring & Beer, 2005). The emphasis may vary at different stages of the disease, and with different people, ultimately aiming toward long-term adaptation to changes in ability, environment and roles, over the life span (Kesselring, 2004). Adaptation is core to many models of rehabilitation, especially those focused on performance and participation (Hagedorn, 2001; Foley, 2008). Adaptation can be both positive and negative, with some naturally occurring adaptations being considered maladaptive. For example, people with cognitive impairment may automatically reduce workload or limit social contacts because of the difficulties of maintaining those activities. For them, this could be a positive adjustment, or it could lead to reduced social contact and withdrawal, a negative adaptation, because of the impact on quality of life (Prigatano, 1999).

Given the lack of curative treatments for MS, pharmacologic therapies are typically provided in conjunction with rehabilitative approaches to managing the effects of MS and resultant disability. The conjunction of physiotherapy and anti-spasticity medications to maximise mobility (DeSouza & Bates, 2004), or occupational therapy in conjunction with anti-fatigue medication for maximising performance in activities of daily living (LaRocca et al., 2006; Finlayson, Garcia & Cho, 2008), are examples of this multi-modal approach.

In rehabilitation, one underlying factor, often assumed, is sufficient cognitive ability to learn and remember, to tolerate the demands of, and contribute sufficient effort to, a rehabilitation programme or activity (Raskin & Mateer 1994). In samples of people with MS, cognitive impairment is negatively related to treatment outcome (Grasso et al., 2005), and more generally, neuropsychological impairment and low mood may contribute to poor coping ability in people with MS (Cox & Julian, 2005).

The factors that predict rehabilitation outcomes in people with MS have been considered to only a limited extent (e.g. Rao, et al., 1991b; Langdon & Thompson, 1999), with sometimes difficult to interpret results. Langdon & Thompson (1999) studied a sample of 35 people with MS who were admitted for 3 weeks of goal-oriented neurorehabilitation. They reported that motor disability, (measured by the Functional Independence Measure, or FIM), verbal intelligence (measured by WAIS-R Vocabulary Test) and cerebellar function (measured by Kurtze Cerebellar Function Subscale), were the best predictors of FIM *Motor* scores on discharge.

One consideration for this finding was that there was little evidence of widespread or severe cognitive impairment and there were ceiling performance issues with the FIM *Cognitive* scale, meaning that statistical associations with outcome were probably difficult to establish (Langdon and Thompson, 1999; Lawton et al., 2006). In general, the sample did not have significant cognitive impairment, perhaps making them more likely to be selected for neurorehabilitation in the first instance; this might be considered a tacit acceptance of the negative prognostic impact of neuropsychological impairment. As the authors consider, they may have ‘studied a highly motivated group’ (Langdon & Thompson, 1999:99). Implicit assessments of motivation and cognitive function may have therefore been clinical judgements made by the admitting team, leading to this study only focusing on the factors that suggest rehabilitation success in terms of motor function in people with MS who are motivated, can take part in goal setting and who do not have contributory impairments in learning and memory.

Another study, demonstrating improvements in dexterity, balance and walking speed after three weeks of inpatient rehabilitation had exclusion criteria of people scoring less than 24 on the Mini Mental State Exam (Vikman et al., 2008). As an assessment of cognitive ability, the MMSE has limited validity in subcortical disease, where the tested domains are often grossly intact (Beatty & Goodkin, 1990; Swirsky-Sacchetti et al., 1992), making the cut-off of cognitive ability difficult to interpret. An interesting finding from the study was the lack of improvement on activities of daily living (ADL) tasks, despite motor improvements, supporting the findings of another study, in which comparisons of people with MS who are classified as cognitively impaired, compared to those who are not, suggest globally poorer functioning in daily life for the cognitively impaired samples (Rao, et al., 1991b). Such ADL performance tends not to be, or is only weakly, correlated with physical disability (Rao, et al., 1991a).

In respect of cognitive impairment and its management in rehabilitation, it has been noted that the:

> *'MS population has received relatively little attention with regard to effective rehabilitation of cognitive impairments. When reviewing texts of health professionals, such as occupational therapists and neuropsychologists, treatment related to cognitive symptoms of people with MS is scarcely mentioned'* (O'Brien et al., 2008:766).

Some frequently used measures of MS-related disability may also be biased towards the physical effects of MS, rather than the cognitive effects (e.g. the EDSS; Richards et al., 2002), either compounding the under-representation of cognitively impaired participants in outcome and rehabilitation evaluation studies, or contributing unknown effects to findings.

In some studies of rehabilitation interventions, cognitive measures could be considered insufficient; such as the use of a single Likert-scored question about difficulties thinking and concentrating (Di Fabio et al., 1998). Both single-item measures of cognitive ability, and the subjective report measurement approach potentially undermine validity, generating results possibly reflecting mood states, as opposed to cognitive performance (Randolph, Arnett & Freske, 2004; Goverover, Chiaravalloti & DeLuca, 2005).

A recent review of the evidence base for cognitive rehabilitation in MS (O'Brien et al., 2008) proposed that the domain was in its infancy, but given the impact of cognitive impairment across the range of human activities, the treatment of these deficits by methods including strategies, compensatory methods and mnemonic approaches was important O'Brien et al., (2008) recommend that for learning and memory, techniques such as a story memory technique or self-generation of to-be-remembered information should be used. In addition to memory-oriented treatment, executive function training, including activities that loaded on strategic thinking, demonstrated improvement in the study, both in relapsing-remitting and secondary progressive MS participants.

An important derivation from this review is that memory rehabilitation may work for people with MS, notably when they focus on the use of strategic methods. Self-generation of to-be-remembered information is proposed to enhance encoding and therefore later recall, addressing both acquisition and retrieval processes (Goverover, Chiaravalotti & DeLuca, 2008). The paradigm requires that in some way the learner organises the to-be-learned information, for example by drawing it, filling in the missing (to-be-remembered) word in a sentence, or generating the steps of a functional task (Mulligan & Lozito, 2004:177). These approaches focus on the generation of the to-be-learned information alone. Findings from O'Brien et al., (2008), suggest that also generating the strategy itself, and not just the to-be-learned material, might be an additional important consideration; certainly it may have more ecological relevance, because strategy selection is often task specific (Delaney et al., 1998; Turley-Ames & Whitfield, 2003)

Overall, evidence from people with MS not only implicates memory ability in such remediation approaches, as would be expected, but also executive function (Canellopoulou & Richardson, 1998). The role of executive function in the use of imagery mnemonics is proposed in both the selection and deployment of appropriate strategies, thereby reflecting metacognitive processes (Canellopoulou & Richardson, 1998).

One difference between Canellopoulou & Richardson's (1998) study and others, also using MS samples, suggesting that self-generation improves memory performance (Chiaravalloti & DeLuca, 2002; Goverover, Chiaravalloti & DeLuca, 2008), was the dual requirement to generate the strategy and its contents in the first study; this is likely to demand more effortful processing. In comparison, self generation for the content for a *given* strategy may be less cognitively demanding, not requiring for example the same attention to a tasks' extrinsic qualities or to an assessment of internal memory strengths and weaknesses. In the studies these differences in given versus generated strategy perhaps give an insight into a continuum that highlights limitations in metacognitive control of memory in MS.

In the Canellopoulou & Richardson (1998) study, self-generation of mediators for a memory strategy (imagery mnemonics) was found not to assist later recall, whereas experimenter-provided imagery increased performance. In two other studies (Goverover, Chiaravalloti & DeLuca, 2008; Chiaravalloti & DeLuca, 2002), the self-generation of target words in sentences and of missing words in task instructions (e.g. making an omelette: 'begin by beating the ____') was the approach taken. Success of the approach in the latter two studies was proposed to relate to depth of encoding factors. Failure in the first was linked to limitations in the executive control of the demands of generating mediators for the strategy. It seems while the samples may have been able to generate specific pieces of to-be-remembered information, and benefit from this process, they may have difficulty generating material for the strategy itself.

These findings may directly address an assumption made in many studies of memory rehabilitation, and rehabilitation in general. How tasks are often highly structured negates the need to successfully apply an organising principle of some sort, as might be required in many everyday memory tasks. The assumption is of a person with cognitive impairment being able to appraise for what memory situations strategies would be required, which strategies would be useful, and being able to generate useful strategy content (Moffat, 1992; Wilson, 2000). Spontaneous use of strategies also implies there is sufficient awareness of its need (Prigatano, 1999). Given the significant strategic processing, in conjunction with the required awareness of task demands, findings that executive abilities are important appears appropriate. Perhaps each of these studies could have benefited from explicitly focusing on how the learner might identify the need for a strategy, what strategies might work given the task and their own cognitive style, and an error-checking mechanism to assess for success in strategy use.

Generally, rehabilitation efforts aimed at optimising learning performance typically benefit from being supported by interventions to improve patients' accurate appraisal and management of their memory, or cognition in general (Naugle & Chelune, 1990; Giacino & Cicerone, 1998; Prigatano, 1999; Toglia & Kirk, 2000). Indeed, awareness of memory difficulties is positively related to outcomes of cognitive rehabilitation in Alzheimer's Disease (Clare et al., 2004), and is the basis for instituting compensatory strategies, which can reduce disability (Wilson & Moffat, 1992; Kennedy, Carney, & Peters, 2003; Dunlosky et al., 2005). It is noted both from clinical experience, and the literature, that those who have limited insight or understanding of their memory disability, tend to have the poorest outcome in the rehabilitation of memory (Wilson & Moffat, 1992; Prigatano, 1999; Prigatano & Kime, 2003). Conversely, for those with poor memory but good insight, focusing on awareness as the intact ability, may be the optimal approach to take in the management of the memory disorder (Nelson et al., 1994; Clare et al, 2004; Kennedy & Yorkston, 2004). In effect, what a person *'does with what they have, is more important than what they have'* (Ylvisaker & Feeney, 2004: cited in Kennedy, 2006).

Awareness therefore, becomes an important issue in the process of self-regulated learning (Ownsworth et al., 2006; Souchay, 2007). An approach to considering awareness in people with brain injury outlined by Toglia & Kirk, (2000) may be a useful way of considering memory monitoring processes. Toglia & Kirk (2000:60), in their review of awareness in people with traumatic brain injury highlight three perspectives; *metacognitive knowledge*, *emergent awareness* and *anticipatory awareness*. The authors propose that metacognitive knowledge relates in part to stored perceptions about ability, including efficacy-related beliefs and related affective states. Emergent awareness, as the name suggests, develops during the experience of being involved in tasks, and anticipatory awareness relates to prior appraisal or development of a conceptual understanding of the task (Toglia & Kirk, 2000).

This underlines, in respect of memory functioning, the importance of being able to accurately appraise memory ability, the task as well as select and use appropriate strategies based on likely memory performance. This ability to take a supervisory stance, or meta-stance, on memory functioning is part of what is termed *Metamemory* (Flavell, 1979). This involves the monitoring and control of memory operation (Nelson & Narens, 1990). Metamemory is viewed as a process of awareness, including knowledge about, monitoring and control of the operation of memory (Nelson & Narens, 1990, 1994; Schneider, 1998; Son & Schwartz, 2002; Van Overschelde, 2008; Koriat et al., 2008).

1.7. Metamemory

1.7.1. Defining Metamemory

The memory-oriented processes of encoding and retrieval were previously discussed in relation to performance of people with MS. This approach to memory characterises ability in terms of the amount of information acquired, retained and retrieved. Such a '*quantity-oriented'* (Koriat & Goldsmith, 1996:491) type of assessment of memory, common in clinical environments, focuses in quantifying amounts of recall as the primary measure of ability (Koriat & Goldsmith, 1996; Koriat, Goldsmith & Pansky, 2000; Delis & Kramer, 2000). The assumption made is that good performance is an index of the functioning of a range of underlying processes, including those related to remembers' conclusions about the correctness of their answers. However, in such a task where the number of items correctly retrieved is the indicator of memory ability, a person could output all possible or potential answers, and still achieve good memory scores, despite a high number of false positive answers (Koriat & Goldsmith, 1996; Delis & Kramer, 2000). Consideration of the number of false positive answers does extend understanding of memory performance, because it reflects that one person sets low thresholds in their assessment of what they know, thus over-reporting (Goldsmith, Koriat & Pansky, 2005). Reporting of high numbers of false positives suggests that there are additional subject-controlled mechanisms involved in retrieval (Petrides, 2000; Koriat & Goldsmith, 1996).

The work of Bartlett (1995) and Loftus (Loftus & Palmer, 1974) propose that memory is impacted, at acquisition, by processes of levelling (simplifying), sharpening some details and assimilating aspects of events to fit with a preconceived view of the world (Bartlett, 1995; Roediger, Bergman & Meade, 2000), and again adjusted at retrieval, for example to fit with questioning style (Loftus & Palmer, 1974). Both monitoring - of incoming material, as well as current knowledge - and control operations - sharpening, assimilating and adjustments at retrieval - are therefore implicated. The view of memory as a faithful store of events or information, with each retrieval being the same copy of the original event, has been undermined both by studies of false memory (Mazzoni, 2008), and memory distortions (Roediger & McDermott, 2000).

Accuracy-oriented approaches to memory focus on the correspondence between how accurate memory is in representing a previous event and the event itself. This might be the concern when interviewing an eye-witness (Perfect, 2002; Koriat & Goldsmith, 1996). In this context, the importance of accuracy - *nothing but the truth* - rather than quantity of remembered information is of more concern. The research literature has therefore explored such phenomena as confidence in remembered information, accuracy of what is recalled, and the relationship between the two (Perfect, 2002; Dunlosky & Metcalfe, 2009). This correspondence view of memory focuses on processes such as ‘strategic control’ in memory (Koriat & Goldsmith, 1996:492), meaning that both monitoring of the recalled information to decide if it is indeed accurate, and a decision whether to report it as known, can be made by the rememberer in order to maximise accuracy.

In everyday memory situations, there is an opportunity to calibrate the level of generality of responses, from precise to more ‘coarse’, to maintain levels of accuracy (Goldsmith, Koriat & Pansky, 2005). Such strategic regulation can be attenuated by asking people to report exactly what they had originally learned, suggesting that where such subject-controlled processes are limited by the report option, recall performance is akin to a normal forgetting curve (Goldsmith, Koriat & Pansky, 2005). This constraint on the use of person-controlled performance underlines some of the limitations of quantity-oriented approaches to characterising memory ability, and points towards a range of variables that might be both investigated and manipulated in considering accuracy-oriented approaches. Having this accurate appreciation of ones' own memory performance, the factors that influence it, and qualities of the to-be learned information (e.g. its complexity), are important in learning (Thiede, Anderson & Therriault, 2003).

From a historical perspective, only since the 1960’s were these qualitative aspects of memory phenomenon approached in an experimental sense (Hart, 1965; Flavell, Friedrichs & Hoyt, 1970; Flavell, 1979; Metcalfe, 2000; Dunlosky & Bjork, 2008a), around the time that Tulving & Madigan (1970) suggested that this avenue of research could be an important avenue for the development of understanding of the psychology of memory. Up to the early 1990’s there were disparate research directions; in education, cognitive psychology and, less commonly, in neuropsychology (e.g. Hart, 1965; Shimamura & Squire, 1986; Janowsky, Shimamura & Squire, 1989; Pannu & Kaszniak, 2005). After publication of a proposal for formalising research in the area and related measurement considerations (Nelson & Narens, 1980, 1990; Nelson, 1984, 1996a,b) a framework became available to help unify the approach to research with a view, more recently, towards application (Dunlosky & Bjork, 2008; Perfect & Schwartz, 2002). A more contemporary development has been an attempt to link models of metamemory, to cognitive neuroscience (Shimamura, 2000, 2008; Schwartz & Bacon, 2008), life-span development and neuropsychological impairment (Beatty & Monson, 1991; Moulin, 2002;

Hertzog, 2002; Pannu & Kaszniak, 2005; Perrotin et al., 2006; Souchay, 2007; Souchay et al., 2007; Dunlosky & Metcalfe, 2009).

The relationship of metamemory to learning and memory has been investigated extensively in educational settings, in topics such as the regulation of study time (Thiede, Anderson & Therriault, 2003), or decisions about when to terminate learning (Maki & McGuire, 2002). Both allocation of study time and termination of study are classified as control operations because they involve modulation of cognitive processes to optimise learning (Nelson & Narens, 1990; Dunlosky et al, 2007; Hacker, Bol, Keener, 2008) It is proposed that control operations such as these are based on the results of monitoring of ongoing learning experience or of the features of the to-be learned material (Nelson & Narens, 1990; Metcalfe & Finn, 2008).

In summary, metamemory therefore consists of accurate knowledge and beliefs about memory performance (‘I am good at remembering names’) and abilities related to monitoring and control of memory performance (Nelson & Narens 1990, 1994; Scarrabelotti & Carroll 1999; Mazzoni & Kirsch, 2002; Perfect, 2002; Souchay & Isingrini, 2004; Dunlosky & Bjork, 2008). Additional abilities in appraising tasks to anticipate memory demands are also implied (Nelson & Narens, 1990; Metcalfe & Finn, 2008). Metamemory, has also been demonstrated to be selectively impaired in people with similar levels of memory impairment (Shimamura & Squire 1986), suggesting some independence from memory functions and some reliance on other domains of cognition (Pannu & Kaszniak, 2005; Schwartz & Bacon, 2008)

Metamemory belongs to a broader class of operations relating to cognitive functioning, *Metacognition.* Metacognition attempts to embody the implementation of self-appraisal and management processes in respect of one’s individual cognitive performance (Metcalfe, 2008; Dunlosky & Metcalfe, 2009). The relationship of this meta-level to its respective cognitive process is presented by Nelson & Narens (1990) in three principles.

First, there are cognitive processes occurring on at least two levels simultaneously – the object level (actual cognitive processing) and the meta-level. Second, that the meta-level contains a ‘dynamic model’ of the object level (Nelson et al., 1999). Third, two relationships exist between the object and meta-levels of cognition – those of monitoring and control. A diagrammatic representation, given by Nelson & Narens, (1990:126) is presented below.

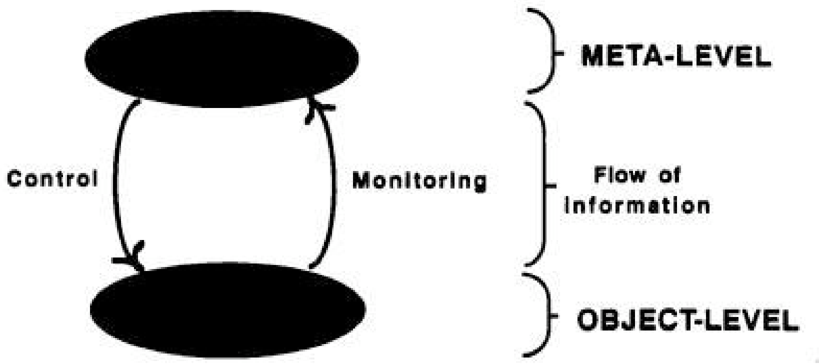


Figure 1.2 **Nelson & Narens' (1990) summary of the relationship between object level and meta-level processes.** Nelson, T. & Narens, L. (1990) 'Metamemory: A theoretical framework and new findings', *The Psychology of Learning and Motivation,* 26 pp.125-173

Nelson suggests that control processes allow for modification of the object level functions, for example to '*terminate, initiate* or *continue but modify'* behaviour (Nelson et al., 1999:74). The control process itself does not obtain information from the object level. The monitoring component acts as the detection mechanism and has been likened to eavesdropping on a telephone conversation (Van Overschelde, 2008), where awareness, but not ability to influence, is possible. Van Overschelde (2008) goes on to propose however that the monitoring process is likely to be more active than the eavesdropping simile suggests, as there are likely to be goal-oriented judgments behind the monitoring. Rather than just through the object level 'informing' the meta-level, this monitoring is achieved by selection of relevant cues to assist monitoring accuracy. Such cues are likely to depend on the judgment task. This more active construction of the monitoring process has implications for considering the metamemory task, as well as the judgment maker.

Nelson & Narens' (1990) model of the relationship of task-based metamemory monitoring and control processes, and how they might be associated with acquisition and retrieval stages of memory, adapted by Dunlosky, Serra & Baker (2007), is presented in figure 1.3 below:

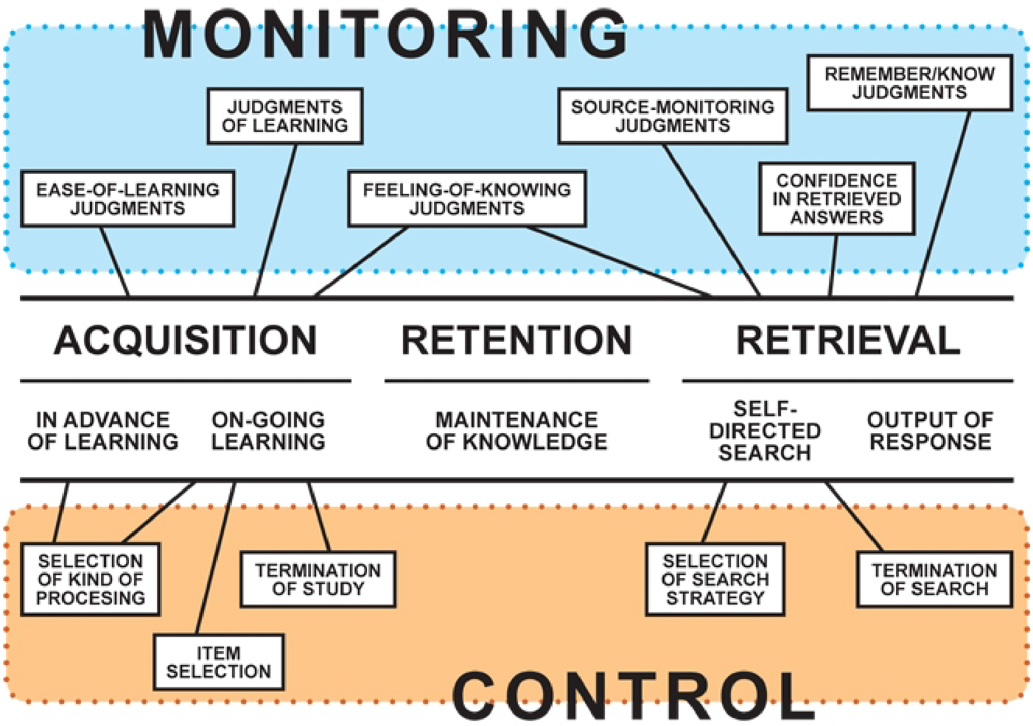


Figure 1.3 Summary of metamemory judgments and their associations with memory processes, adapted by Dunlosky, Serra & Baker (2007) and presented in Durso, F Nickerson, R., Dumais, S., Lewandowsky, S & Perfect, T., Eds., (2007) *Handbook of Applied Cognition* 2nd Ed., Wiley, New York.

Monitoring-related processes comprise a sizable component of metamemory research (e.g. as reviewed by Koriat, 2002b) spawning a range of indices of monitoring-related processing: Feeling of Knowing, Judgment of Ease of Learning, Judgment of Learning, Retrospective Confidence, Source Monitoring (Nelson & Dunlosky, 1991; Dunlosky, Serra & Baker, 2007; Dunlosky & Bjork, 2008a; Narens, Nelson & Scheck, 2008). This focus on monitoring-related processes is notable outside of the domain of developmental psychology (Koriat, 2002b). Clinically, in supporting increased awareness of memory performance, as an important part of rehabilitation of memory deficit, monitoring is a key area for investigation (Wilson & Moffat, 1992; Prigatano, 1999; Prigatano & Kime, 2003; Clare et al., 2004; Clare, 2004). As proposed by Nelson & Narens (1990), and supported in recent studies (Metcalfe & Finn, 2008), monitoring processes are the method by which control operations are instituted, which might suggest primacy in terms of rehabilitation intervention (Moulin, 2002; Dunlosky, Kubat-Silman & Hertzog, 2003; Dunlosky et al., 2005). The following review will therefore focus on aspects of the monitoring processes within metamemory.

1.7.2. Metamemory Monitoring

Metamemory monitoring processes are broadly considered to be of two types; declarative and procedural. Declarative metamemory is considered to be the 'factual knowledge about' memory (Schneider 1998:1), beliefs about memory competency (Koriat et al., 2008) to which some add the ability to 'explain decisions relating to memory actions' (Schneider & Lockl, 2008:392). This is akin to what Koriat and others (Koriat & Levy-Sadot, 1999; Koriat, 2007; Koriat et al., 2008) have termed information-based metacognitive judgement, in that the appraisal is based on 'information' about memory known from memory use and beliefs, distilled to a perspective on one's memory.

The alternative to declarative or information-based metamemory, termed *procedural metamemory*, embodies an awareness of the proceeding of memory performance, which relies in part *'on mnemonic cues that are devoid of declarative content….* [these] *cues are derived from the very experience of learning, remembering and deciding, rather than from the content of thought'* (Koriat et al., 2008:119). Also termed experience-based (Koriat & Levy-Sadot, 1999; Koriat, 2007; Koriat et al. 2008), these metamemory judgments therefore implicate monitoring processes during task performance (i.e. 'on-line' monitoring). Examples of these 'content-less cues' include retrieval fluency (Koriat & Levi-Sadot, 1999) that experientially might tell the rememberer that speed of retrieval relates to knowing, and therefore accuracy.

Information-based and experienced-based metacognition differ in terms of the data used to make the judgment (Koriat et al., 2008). Information-based judgments tending towards reliance on knowledge which could be affect-based causal attributions- 'I am no good at memory tasks' (Randolph, Arnett & Freske, 2004), expectation based - 'I should know the answer to this' (Glenberg & Epstein, 1987) or culture-based beliefs - 'I am old and memory declines as you get old, so I will forget this' (McFarland, Ross & Giltrow, 1992; Hertzog & Hultsch, 2000; Hertzog, 2002). Importantly, Koriat et al. (2008) suggest that these attributions should impact on information-based judgments. Information-based judgments are considered more global in their focus, whereas experience-based processes are borne out of the 'on-line' processing during the stream of activity, and are subjective experiences (Koriat et al., 2008). This might mean that more global judgments, not based on experiential memory processing, might be impacted more by stored perceptions about one's memory, and related affective and efficacy factors (Toglia & Kirk, 2000). One common measure of information-based global memory judgment would be a memory questionnaire (Gilewski, Zelinski & Schaie, 1990; Hertzog, 2002; Zelinski & Gilewski, 2004); another might be a clinician's questioning about a patient's memory.

Experience-based judgements are proposed to have multidimensional foundations (Nelson, Gerler & Narens, 1984; Krinsky & Nelson, 1985; Nelson & Narens, 1990) considered to be made based on cues from the experience of learning and remembering, including feelings of familiarity with the question (Reder & Ritter, 1992), or accessibility of even partial information about the target item, or fluency experiences during recall (Dunlosky & Nelson, 1992; Benjamin Bjork & Schwartz, 1998; Son & Schwartz, 2002; Koriat et al., 2008). Generally these are considered heuristic approaches to monitoring, in that they are used a mental rules of thumb to guide judgments about memory ability (Dunlosky & Metcalfe, 2009).

The processes underlying information-based and experience-based metamemory also have some similarities, according to Koriat et al. (2008). Inferential reasoning, as it relates to information-based judgments might be explicit reasoning relating to the judgment to be made, whereas in experience-based judgments, *'various mnemonic cues are used en-masse to give rise to a sheer intuitive feeling'* (Koriat et al., 2008:120); experience-based process in this sense are considered to be '*parasitic'* on object-level memory processes (Koriat et al., 2008). Memory experience cues may imply a level of direct access to target items in memory that help with the metacognitive judgement. For example, good (fast or sizable) access will support a high confidence. Weaker indications of availability will provide perhaps more ambivalent judgments of confidence. Finally, tip-of-the-tongue experiences, where confidence is high even in the absence of access, suggest a role for a separate contribution of the familiarity sensations (Son & Schwartz, 2002).

The distinction between information and experience-based contributions to memory monitoring is potentially useful in thinking about structuring research, assessment and interventions into metamemory, though not without some consideration of how they might be differentiated in practice. If one such inferential signal, cue familiarity, is considered, some blurring of the information versus experience-based contributions can be seen (Schwartz & Metcalfe, 1992; Metcalfe, Schwartz & Joaquim, 1993). Cue-familiarity suggests that when judgment about a question in the sphere of *Current Affairs*, someone might infer they could trust a guess as an answer and if they consider themselves ‘good’ at current affairs. The cue (e.g. a particular question) might be familiar because of repeated exposure, perhaps during the development of expertise in this domain. This, according to Koriat et al., (2008) would be considered an explicit, information-based inference. This is knowledge about memory strength in this area of semantic memory. Alternatively, some view cue-familiarity as an experience-based resource for judgment making, a heuristic that familiarity experiences generated by a cue suggest access to its target. Dunlosky & Metcalfe (2009) suggest one approach to defining the difference between the two as considering one *domain-familiarity* the other *cue-familiarity*; the former being information-based, the latter experience-based.

Notable by its absence in this review is consideration of direct-access to the contents of memory, which none of the above assume necessary for the judgment making. The role of direct-access will be considered in discussing specific metamemorial judgments, but for the present it can be summarised as not being typically given enormous weight in some judgments (Koriat, 2002) and this may be for two reasons. Firstly, a focus on direct-access perhaps removes the *meta* from metamemory (Dunlosky & Metcalfe, 2009; Kimball & Metcalfe, 2003). Second, there is evidence that judgments in some circumstances are made faster than could retrieval (Reder & Ritter, 1992). Son & Schwartz (2002) suggest that direct-access might also be a graded resource, including partial access to the target information, which would guide a judgment being made about likely accessibility of the required information (Leonesio, 2008; Metcalfe, Schwartz & Joaquim, 1993)

In summary, interpreting the literature on how metamnemonic judgments are supported, three areas of consideration are indicated; mnemonic, because of its contribution to experience-based judgments, executive, because of the inferential nature of many monitoring judgments and memory-efficacy in how it relates to knowledge-based contributions (Beatty & Monson, 1991; Koriat & Levy-Sadot, 1999; Johnson et al., 2000; Koriat, 2007; Van Overschelde, 2008; Koriat et al., 2008). This latter element is more difficult to operationalise as it can be a function of culture, beliefs, attitudes or distortions in self-appraisal related to thinking style; the latter, as might be found in mood disorder, making it a final relevant factor in accuracy of memory judgment (Elliot & Lachman, 1989; Clark, Beck & Alford, 1999; Hertzog, 2002; Koriat, 2002).

Therefore, while memory tasks are an appropriate and necessary way of assessing metamemory, consideration of the memory abilities of the judgment-makers is necessary, especially in those with memory impairment. One assumption might be that with impaired object-level memory ability, memory-experience cues themselves may be degraded, reducing their diagnostic utility for judgment makers. In respect of the other processes highlighted, if inferential abilities were impaired, both the selection and evaluation of appropriate cues to optimise judgment accuracy might also be limited.

1.8. Monitoring Judgements

Nelson suggests that '*the primary* [researcher] *tool for generating data about metacognitive monitoring is the person's subjective report'* (Nelson et al. 1999:74). Nelson & Narens (1990) underline that the introspection cannot be assumed to be a perfectly valid reflection of the processes of memory, but instead, accurate and distorted products of introspection are the data of interest, and to be explained (Nelson & Narens, 1990; Nelson et al., 1999). Despite the distortions that occur, this system of introspection is used to guide memory-related goals and behaviours (Van Overschelde, 2008).

Monitoring mechanisms, with their distortions, are potential targets for training and adjustment, or de-biasing (e.g. by feedback schedules, Dunlosky & Metcalfe, 2009) within the rehabilitation environment, and are therefore of interest. In developing interventions for a clinically relevant distortion however, factors that contribute to distorted judgments are of initial interest to clinicians, as they have been to educators, because they may themselves be fruitful targets for intervention (Randolph, Arnett & Freske, 2004; Julian, Merluzzi & Mohr, 2007). In advance of considering performance of neurologically impaired samples, two aspects of metamemory research need consideration - the measures frequently used, and the methods of measurement for these judgments. Three common judgments are discussed, with a specific focus on the proposed mechanisms involved in making each judgment.

In a memory task, a number of stages can be considered - encoding and retrieval being the most emphasised aspects of memory performance in the clinical environment. In the proceeding of these processes, Judgments of Learning can be made during and after learning, Retrospective Confidence Judgments after retrieval and for non-recalled items, Feelings of Knowing can be made. Each of these judgments is now discussed with a view to understanding the roles of different contributions, so as to begin to outline the relationships between judgments and contributing factors in the sample used for this study.

1.8.1. A Judgment of Learning

A Judgment of Learning (JoL) is often made during or after an initial period of study and the judgement is of likely future recallability (Nelson & Narens, 1990). In a list-learning task, such as a shopping list, a judgment of learning might be made after generating and writing down the list. If the judgment suggested that the list had been committed to memory, it might be thrown away instantly and a trip to the supermarket embarked upon. Alternatively, if the judgment suggested the list was not known it might lead to a second learning episode and a quick self test of items. A number of features of Judgment of Learning can be revealed in this event.

Nelson & Narens (1990) suggest that the basis of the judgment is generally in an assessment of currently recallable items, using covert retrieval. From this suggestion two approaches to testing grew - accuracy of immediate JoLs, which are made immediately after learning of information, based on an assessment of currently available target items in memory. These relate to an assessment of short-term memory stores.

Also investigated are delayed JoLs, made after a delay and presumed to be based on and assessment of availability within long-term storage (Nelson & Dunlosky, 1991; Narens, Nelson & Scheck, 2008). In the example given above, an assessment of having learned a shopping list made *after a delay* is likely to be more accurate because the judgment is based on interrogating long-term memory stores, the same memory store from which the list will be drawn on arrival at the supermarket. Therefore delayed Judgments of Learning are proposed to be better predictors of actual future recallability. The differences between accuracy of immediate and delayed Judgments of Learning were explained on the basis of a *Monitoring Dual Memories* hypothesis (Nelson & Dunlosky, 1991).

Nelson and Dunlosky (1991) note differences in accuracy of the judgment in predicting future recall. If made immediately, JoLs can be inaccurate (gamma correlation between predicted and actual recall was $G = 0.38$). Accuracy was high in delayed JoLs ($G = 0.90$). This was explained by the proposal that people were monitoring from within two different memory systems for the immediate judgment. In making an immediate JoL, monitoring happens from the recently given information as well as long term memory. Long term monitoring is required to assess learning, whereas short term only current availability. This dual monitoring in immediate JoL leads to interference; concurrent short-term availability-monitoring interfering with monitoring from the destination memory store for the target, in long-term memory (Dunlosky & Metcalfe, 2009).

For delayed JoLs monitoring was proposed to be concordant, in that monitoring was only of long-term store. Given that this memory store is the same store from which tested retrieval would take place, accuracy would increase with judgment at this delayed stage. This hypothesis was used to explain the delayed Judgment of Learning effect - the increase in predictive accuracy of the judgment noted after delay.

Typical JoL tasks generate the gamma correlations based on a series of target item predictions, subsequently compared to tested performance for that item, providing an index of relative accuracy. Relative accuracy gives an assessment of each items' judged learning compared to that of all other items. In some studies the JoL is made at a more global level, on groups of items (Hacker et al., 2000); the shopping-list example would be an example of the judgment being made on a group of items. In other versions of JoL tasks, repeated presentation of target items is given, with JoLs made between each presentation (Meeter & Nelson, 2003).

Notable with the repeated presentation studies is a second characteristic of JoLs; what is termed the *underconfidence with practice effect* (Koriat, 2002). This is noted where predicted performance increases in smaller amounts that actual performance, leading to an increasing disparity (underconfidence) over repetitions of target learning (Koriat, 2002; Dunlosky & Metcalfe, 2009). Returning to the shopping list example, repeated reading followed by a JoL would be expected to lead to increasing underconfidence in how many items on the list would be remembered.

One explanation for this is that people use their performance on the previous test to estimate future performance, so that predictions lag behind performance - an example of an inferential mechanism contributing to the judgment, and termed the *memory for past test heuristic* (Finn & Metcalfe, 2007). Of interest, this underconfidence with practice is typically seen to be abolished if the judgment is made after a delay, which perhaps supports some differences between the monitoring processes in immediate and delayed JoLs

Debate continues over what processes or mechanisms contribute to JoL accuracy; that delays produce more accurate results is a robust finding across a range of studies (Koriat, Sheffer & Ma'ayan, 2002; Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009) There also continues some debate as to whether this judgment reflects mnemonic phenomena as opposed to a metamemory processes (Kimball & Metcalfe, 2003), though it is likely that some assessment of memory experience, and perhaps inferential judgments such as memory for past test do mould the predicted performance component of the judgment. On balance however, the mechanisms associated with accurate performance do appear primarily relate to memory and memory experiences, rather than inferential judgments (Spellman & Bjork, 1992; Kimball & Metcalfe, 2003; Narens, Nelson & Scheck, 2008)

1.8.2. A Retrospective Confidence Judgment

Retrospective Confidence Judgments (RCJ) are judgments of the likelihood of being correct, made after recall (Dunlosky & Metcalfe, 2009). One debate, given the judgment is based on the products of retrieval, is the extent to which so-called 'privileged access' informs the judgment. Privileged access relates to the access that people have to internal contents of mind (Nelson et al., 1986), and this contributes to inferences about performance judgment. Typically this is considered to relate to experiences of fluency or ease of processing, even if they are devoid of content (Koriat et al., 2008). However, given these experiences are considered *parasitic* on object level performance (Koriat et al., 2008), some association with object level memory performance is likely. Alternative sources of information that do not relate to privileged access include assessments about the difficulty of a task, an extrinsic factor that may mediate judgment (Maki, 2008).

In studies of RCJ, both confidence judgments about performance, and known item difficulty have been shown to predict performance in general knowledge tests, suggesting there may be some sensitivity to objective question difficulty (Maki, 2008). These studies went on to investigate judgments made on newly learned material to find the best predictor of performance; confidence or known item difficulty. The key finding was that confidence judgment was a better predictor of performance, notable in more difficult learning tasks, supporting a view that idiosyncratic mnemonic information (privileged access) may be the key resources used, compared to test difficulty (Maki, 2008).

To investigate the contribution of information-based and experience-based processes to confidence judgments, Koriat et al., (2008) presented a general knowledge task to high-school students. A key manipulation was allowing either free-report of as many reasons for their answer, or a forced-report of a specific number of reasons (matched to the number given by the free-report participant). After reporting reasons under each condition, a confidence judgment was made. Their hypothesis was that ease of retrieval was a key cue used in informing confidence. In the free-report condition, ease of retrieval and number of reasons given was supported as having a positive relationship with confidence judgment accuracy. In the forced report, the effort at reporting a number of reasons was proposed to give an effortful retrieval experience, leading to subjective ratings of lower confidence. While these findings relate mainly to semantic memory, the new learning studies by Maki (2008) support the idea of mnemonic cues, such a fluency and ease, as having an important role in confidence judgments about newly learned text material also.

One other factor investigated in confidence judgments has been that of the contribution of individual differences, including mood disorder, to under or over confidence. Perfect (2002) reports a number of studies (looking at personality factors that might impact on confidence, across both semantic (general knowledge) and episodic tasks (eye witness scenarios), with the conclusion that neuroticism and social desirability may be relevant factors, but generally, across a range of such measures there was little support for the relationship between measured personality factors and accuracy in either memory tasks.

A related question is of the relevance of mood disorder, investigated by Fu et al (2005) from two standpoints. First, that mood disorder (in this study major depressive disorder) confers ‘depressive realism’ which might be expected to generate more accuracy. Alternatively mood disorder might create a negative bias, leading people to underestimate their abilities, so being more underconfident that their non-depressed peers. While not using memory tasks alone, Fu et al., (2005) concluded that depressive realism was not a feature in confidence judgments. The alternative hypothesis of underconfidence was not supported.

In summary, RCJs are supported to have their basis in the various cues that might accompany memory experience; fluency and ease of retrieval being two proposed indicators. Mood and personality factors do not appear to bias the judgments. Task structure and memory type (episodic or semantic) may relate to how the judgment is made because people probably have some ideas about their domain expertise in semantic memory tasks; this domain expertise might be considered an awareness of task difficulty. When awareness of task difficulty was compared to memory-experience to establish which was the primary cue associated with confidence, mnemonic cues were supported as the key factor.

In assessing contributions to accuracy in this study therefore, memory ability is likely to be supported as the primary contribution. However, given that memory experience, as opposed to memory content, seems to be the key issue, and assuming that this experience may be available even with absent recall, it might be expected that accuracy in this judgment could be expected to be high, regardless of tested memory ability. Experiencing *no fluency* in a retrieval context, might be just as useful a mnemonic cue to accurately estimating confidence as low.

1.8.3. A Feeling of Knowing Judgment

On a memory task there are typically two outcomes - items correctly recalled, and items incorrectly or not recalled (wrong and 'don't know'). For those items that participants are unable to provide correct recall, a Feeling of Knowing judgement can be made, followed by a recognition trial to see if a previously unrecalled target can be recognised. A Feeling of Knowing judgment is a rating (high to low) made for such unrecalled items, of the chances that it would be recognised if seen. Feeling of Knowing (FoK) therefore seeks to assess how accurate judgments, not based on direct access to memory contents, are in predicting future recognition.

Comparison of the strength of the Feeling of Knowing and subsequent recognition performance forms the basis of Feeling of Knowing indices (Nelson, 1996; Nelson et al 1999; Shimamura, 2000). The *judgement* is in participants assessing the likelihood that the currently unrecallable item will be recognised. Dunlosky & Metcalfe (2009) provide a review of the Feelings of Knowing, concluding that the function of the FoK is to inform retrieval strategy if it indicates a criterion level of knowing, or to terminate retrieval attempt, if it does not indicate knowing.

The first experimental procedure for testing the accuracy of FoKs was published by Hart (1965), and focused on FoK accuracy. Hart states: '… *asking how FoK memories are retrieved presupposes that the FoK experience is an accurate indicator of what is in memory'* (Hart, 1965: 208). The study involves a procedure structured as a 'recall-judge-recognise' task (Hart, 1965:209), wherein participants were given a test of semantic memory and for those items that they could not remember, a judgment about their feeling of knowing the correct answer was taken. The FoK was measured by the participant rating their FoK as 'Yes' or 'No' in the first experiment, and on a scale of 6 down to 1 in the second, with 6 being the highest FoK. This was followed by a recognition task in which the correct answer and three foils were given; the participant was asked to select their chosen answer for all questions, regardless of whether they had originally answered correctly or not. Only data for the original 'don't knows' was analysed in examining the accuracy of the FoK, wrong answers were excluded. From two experimental manipulations, Hart suggested that FoKs were accurate in predicting subsequent recognition, especially at the extremes of highest FoK and lowest FoK ratings. Further, he proposed that accuracy was an indication of a participant knowing what was, and was not, in memory store, without having access to it.

Two interpretative points worthy of note are the assumption made by Hart that there was a basis of the FoK in access to memory contents - so called direct-access, so as to make an appraisal of 'target strength' (Dunlosky & Metcalfe, 2009); that is, they indicate, according to Hart, the state of memory storage of items. Additionally, the interpretation of differences in accuracy in the low FoK conditions between experiments I and II is worthy of consideration. There was little concordance between low FoK and incorrect performance on recognition trials in experiment I. Hart (1965) explained this to be because the foils on the recognition trial were not being sufficiently related to the target, making it too easy to pick the right answer. When foils were made more related in the second experiment, there was more evidence of concordance between lowest FoK and actual performance (i.e. not being correct).

Hart explains the decision to change foils on the basis of participants using inferences and domain knowledge to lead them to correct answers on low FoK items, reducing their accuracy by them getting the low FoK items correct. This is perhaps ironic given his proposal that the FoK judgment is based on a direct-access mechanism, rather than an inferential one. These two issues - the direct access to memory and the attempt to limit inference-based selection of recognition target, both mirror subsequent debates about the basis of FoK judgments. More recent analyses of FoK judgments support inferential rather than direct-access accounts (Dunlosky & Metcalfe, 2009).

Importantly, the first step in a FoK procedure is to exclude the remembered material; this also, in part excludes the most obvious direct access mechanism of retrieval. Thereafter a number of types of processing have been proposed, under differing paradigms. Broadly the proposed contributing effects are first, inferential in that they relate to the use of heuristics about relevant factors surrounding the target, such as perceived expertise in the domain. Second, perceived ‘target-accessibility’, or being able to retrieve an aspect of the target, provides support to a feeling of eventually being able to recognise the target in the recognition trial. One example of this might be familiarity with the question, or cue-familiarity (Leonesio, 2008; Dunlosky & Metcalfe, 2009). Of some import here is that many of these understandings about Feeling of Knowing judgments relate to semantic memory, where domain knowledge may be relevant.

Leonesio (2008) draws some differences between others in his use of terms - using *remembrance* to indicate some of the target accessibility type information, such as memory of the encoding context, memory of the cue. There have been some attempts to separate these two heuristic approaches to FoK judgments, but in considering, for example, cue-familiarity either heuristic could explain the process. Cues could be familiar because of expertise in a domain of general knowledge about authors, meaning that a cue in a test of authors would be familiar - this could lead to a reasonable inference that one is likely to be able to recognise an author or could lead to perceived target accessibility, because of familiarity with an aspect of question itself.

One potentially important finding in respect of these heuristic approaches to the FoK judgment is that, while both may be used, they may be deployed differently, depending on the task. This may relate to the level of expertise of the judgment maker; with limited expertise, heuristics may be invalid (Perfect & Hollins, 1996, Perfect, 2002). It may also relate to how the FoK task has failed. According to Koriat & Levy-Sadot (2001), if cue-assessment does not trigger sufficient familiarity, then the target accessibility search may be aborted; in this sense there may be some hierarchical organisation of cue-based heuristics. This proposal may fit well with findings by Reder & Ritter (1992) that FoK judgments can be made faster than actual retrieval would, suggesting they are not based on retrieval availability (direct access), but on cue familiarity. Nelson's (1984:297) use of the term 'memory for prior encounters' as being relevant to FoK judgments, could fit well with this proposal. Depending on the strength of that cue-familiarity, further processing may continue or be terminated.

In summary, the Feeling of Knowing judgment, perhaps unlike Judgment of Learning and Retrospective Confidence Judgments is proposed to relate to inferences made on the basis of more scant cues from memory experience, in part because recall has failed when this judgment is being made. It seems plausible therefore that the judgment may harvest cues from a range of potential informants, from which an inference about likely recognition will be made. Expectations for the factors that contribute are twofold; executive abilities are implicated in the inferential aspects of the judgment and mnemonic perhaps in respect to the range of distal cues from which those inferences would be made.

1.8.4. Measurement Approaches in Metamemory

Aside from the investigation of factors contributing to metamemory judgments, a number of other debates continue with regard to measurement paradigms; how best to calculate accuracy (Hart, 1965; Nelson, 1984; Nelson & Narens, 1990; Schraw, 1995; Wright, 1996; Nelson, 1996; Nietfeld, Enders & Schraw, 2006), what is accuracy (relative or absolute; Nelson, 1984; Koriat, 2002; Nietfeld, Enders & Schraw, 2006) and what stimulus materials to use - previously learned knowledge or new learning (Nelson & Narens, 1980; Janowsky, Shimamura & Squire, 1989; Nelson & Narens, 1990; Souchay, Isingrini & Espagnet, 2000).

Calculation of the association between predicted and actual performance is typically by means of a number of measures including both parametric and non-parametric correlations (Goodman & Kruskal 1954; Nelson, 1984; Schraw, 1995; Nelson 1996b), difference scores (Hart, 1965), proportions correct or sign of difference scores and bias scores

(Nelson, 1984). In principle, accuracy is described by calculating the relationship between 'hits' and 'misses' between judgment and performance. Concordant responses (hits) are those in which high judgment strength associates with successful performance, or low judgment strength with a failed performance (Hart, 1965; Nelson & Narens, 1990; Nelson, 1996). Both are considered concordant because the strength of judgment is predictive of performance. Disconcordant responses (misses) relate to predicting success, but failing or predicting failure, but succeeding. Accuracy levels for these judgments vary with the judgment task and the type of memory task (semantic versus episodic); in general near perfect accuracy is not the norm. For some judgments, above-chance levels might be considered acceptable (Maril et al., 2003; Leonesio, 2008).

1.8.5. Memory Self-efficacy Judgments.

Memory-self efficacy (Zelinski & Gilewski, 2004) is typically assessed using memory questionnaires (Nelson et al. 1999), which aim to probe knowledge of memory ability, lapses, (e.g. kind, frequency and severity) or use of strategies. Such self-report measures are mainly evident in studies of older populations (Hertzog, 2002), perhaps because of the range of work looking at how implicit theories about memory and ageing impact on the reporting memory related behaviours in older people (Devolder & Pressley, 1992). Perhaps another reason is clinical relevance, given that, aside from older people, few, without definable neurological changes, present with complaints of memory deterioration. The focus of research in the area has increasingly been on its validity (Herrmann, 1982; Hertzog, 2002).

Questionnaires about memory abilities may have demonstrated internal validity and consistent factor structure (Gilewski & Zelinski, 1988; Gilewski, Zelinski & Schaie, 1990; Hertzog, 2002; Zelinski & Gilewski, 2004) but they may have limited predictive validity (Dixon & Hultsch, 1983; Hertzog, 2002; Zelinski & Gilewski, 2004). Hertzog argues that while the tools themselves may consistently measure something, it may not be report of memory performance. As has been indicated, beliefs (implicit beliefs about memory or cultural stereotypes; Dunlosky & Metcalfe, 2009), mood or personality traits (Zelinski & Gilewski, 2004) may all affect memory perception, rather than recall of specific memory performance instances across a range of everyday tasks (Hertzog, 2002). In testing this theory, Hertzog suggests that self-report could be made more accurate by emphasising the task-specificity of the question.

The impact of lower memory efficacy is that people may avoid memory-loaded activities (West & Berry, 1994) or be less likely to use strategies to aid memory performance (Lachman & Andreoletti, 2006). A final consideration in using questionnaires as an assessment of memory efficacy is that if people do have memory difficulties, they may adjust their activity or environmental demands as a management strategy, and so face fewer instances of memory failure. This might increase their efficacy, leading to the measure being less reliable indicators of objective memory function (Rabbitt et al., 1997). In clinical environments, an analogue of the memory questionnaire might be general questioning about memory abilities, and so it remains relevant to consider its validity as a form of assessment of memory function.

In summary, metamemory is the ability to take a meta-stance on, appraise, and manage memory performance. In respect of monitoring, this stance can be taken in making a global appraisal of memory ability (‘off-line’ judgment), or in making specific judgments about performance in a specific memory task (‘on-line’ judgment). The former, global appraisal of memory is often measured through questionnaires, potentially collecting efficacy judgments about memory, rather than indicators of actual memory ability. The latter, through judgments made in relation to two of the key memory processes - acquisition and retrieval. To a greater or lesser extent then, all metamemory appraisals are considered to contain an element of distortion because they are introspective and subjective. A key component of their assessment is the inclusion of some objective measure of performance against which they can be calibrated (Bahrick, 2008). The nature or level of distortion is therefore a relevant focus in considering the self-regulation of learning, because of the proposed relationships between accurate monitoring of performance and that learning. Of specific interest, are conditions in which additional distortion might be conferred by impairments associated with neurological disease. There is a limited literature base regarding metamemory in neurological populations generally, and in Multiple Sclerosis specifically.

1.9. Metamemory in Neurological Populations

In a review of metamemory in neurological populations, Pannu & Kaszniak (2005) clarify the types of questions typically asked about metamemory in these populations. Specifically, whether metamemory can be normal in people with memory deficits. This is considered from the point of view of individual pathology, as well as from the perspective of the association of executive function (and/or prefrontal lesions) with metamemory (Shimamura & Squire, 1986; Janowsky, Shimamura & Squire, 1989; Moulin, 2002; Souchay, Isingrini, Pillon & Gil, 2003; Schwartz & Bacon, 2008). The groups best represented in the literature are people with Alzheimer's disease (Moulin, Perfect & Jones, 2000; Moulin 2002; Souchay, Isingrini & Gil 2002; Souchay, 2007), for whom, as clinical samples issues arise with floor effects on recall tests, making some approaches to testing metamemory accuracy difficult. Effectively there is little recall data with which to compare prediction (Moulin, 2002). Approaches in this case have used indices of sensitivity to measure how people adjust their estimations before and after experience with a task to implicate both monitoring and control operations (Moulin, 2002; Ansell & Bucks, 2006; Souchay, 2007). These studies suggest that people with AD can adjust performance, can make correct judgments about item normative difficulty and reflect expected forgetting with delay in their judgments.

With progress of the disease, both floor effects and the relative complexity of the FoK procedure may also impact on getting reliable estimations of predicted performance, though generally it seems that for semantic information, accuracy on FoK tasks tends to be nearer control performance (Lipinska & Bäckman, 1997; Souchay, 2007). An additional consideration is that the mnemonic cues used to make inference-based FoK judgments may themselves be unavailable in episodic tasks, which are typically novel, so that memory experiences do not assist with the FoK, perhaps explaining the lower levels of accuracy, compared to semantic FoK tasks (Souchay, 2007).

In respect to cortical versus subcortical disease, Parkinson's disease patients have, to a more limited extent, been investigated for metamemorial performance. While monitoring performance has been suggested to be generally intact on semantic tasks (Pannu & Kaszniak, 2005), there is some evidence that some control operations may be impaired. Souchay, Isingri & Gil (2006) proposed differences in a free recall task between a Parkinson's group and matched controls related to control, rather than monitoring processes. In a self-paced learning condition, recall was poorer in the Parkinson's group, compared to controls, but not in the experimenter paced learning task, suggesting that decisions about allocation of study time were deficient.

One study, which highlights the dissociated nature of memory and metamemory in clinical populations, is that of Shimamura & Squire (1986), which focused on amnesics with differing aetiologies. Of the sample, only those with Korsakoff's syndrome, characterised by severe amnesia in conjunction with executive dysfunction, were impaired on episodic and semantic FoK tasks. In comparison to comparably amnesic participants (of different causes), the Korsakoff's participants were impaired and the others performed at the same level as the control group. The authors suggest that the frontal/executive deficits could explain the differences between the groups. Pannu & Kaszniak (2005), in their review of metamemory in samples with focal frontal lesions make two important observations; that in such samples there may be little evidence of metamemory deficit unless there is some 'weakening of the memory trace' (Pannu & Kaszniak, 2005:112). Their other observation is that semantic memory tasks may not be sufficiently sensitive or effortful to disclose deficits; that tasks may need to be more difficult.

In discussing future directions of research, Pannu & Kaszniak (2005) item a number of questions; relating different aspects of metamemory to executive function, use of a range of neuropsychological tasks, further investigation of the relationship between deficits in awareness of memory problem and depression, and investigation of the relationship of metamemory inferred from questionnaire and performance on experimental tasks. In reviewing metamemory in MS, and formulating the focus for this study, these recommendations are borne in mind.

1.10. Metamemory in Multiple Sclerosis

Studies focusing on metamemory in Multiple Sclerosis are few (Beatty & Monson, 1991; Richardson & Chan, 1995; Scarrabelotti & Carroll, 1998; Scarrabelotti & Carroll, 1999; Landro, Sletvold & Celius, 2000; Randolph, Arnett & Higginson, 2001; Maor, Olmer, Mozes 2001; Randolph, Arnett & Freske, 2004; Middleton et al., 2006; Julian, Merluzzi & Mohr, 2007). Of these studies, only two have specifically investigated the factors that might contribute to metamemory performance, and one to general cognitive monitoring. A brief summary would suggest that mood, memory and executive functions are all implicated (Beatty & Monson, 1991; Randolph, Arnett & Freske, 2004; Julian, Merluzzi & Mohr, 2007). This generally accords with the contributors discussed earlier in respect of metamemory monitoring - attitudes and beliefs, memory, memory experience and inferential processes. Also concordant are general findings that the mix of associated factors relates to different tasks used (Shimamura & Squire, 1986; Beatty & Monson, 1991; Randolph Arnett & Freske, 2004; Pannu & Kaszniak, 2005).

1.10.1. Memory Self-Report in Multiple Sclerosis

An issue, indicated earlier as a concern for memory complaint, is whether subjective report of memory is valid in predicting memory task performance or whether it is a reflection of mood state or other factors.

In two studies of memory self-report in MS (Randolph, Arnett & Freske, 2004; Julian, Merluzzi & Mohr, 2007), a questionnaire was the mechanism of self-appraisal. Randolph, Arnett and Freske (2004) suggest that dysfunctional attitudes mediated the contribution of both mood and executive function to scoring of items on a questionnaire relating to frequency of memory difficulties while reading. Actual memory performance was tested in this study, but did not correlate with the questionnaire score. The absence of consistent association between memory self-report and objective memory score in the literature has been discussed (Dixon & Hultsch, 1983; Hertzog, 2002; Zelinski & Gilewski, 2004). Probably the more interesting finding is a proposed role of executive function, which will be discussed later.

Julian, Merluzzi & Mohr, (2007) also proposed that depressed mood influenced self-report of cognitive abilities, also measured by questionnaire. Self-report was not a predictor of actual cognitive performance, though self-report did become an accurate predictor, after successful treatment of mood disorder. Accepting that this latter study focused on general ‘cognitive’ abilities, rather than memory specifically, it and the study of Randolph, Arnett & Freske (2004), propose a relationship between mood, and negative attitudes, and memory self-report. This is in line with the evidence base that mood, and other personality characteristics, such as conscientiousness, are associated with this method of self-assessment of memory (Hertzog, 2002; Zelinski & Gilewski, 2004). Additionally, there is a proposed direction, in which depressed mood causes the lower memory ratings. In Julian, Merluzzi & Mohr (2007) causality is proposed through the study design, which was a randomized interventional study. Depression was the independent variable and the accuracy of cognitive function report the dependent. With treated and improved mood, accuracy of report increased.

This latter study design is a traditional approach to the issue of causality, meeting the criteria of precedence in time, non-spurious relationship between predictor and criterion, and a correctly specified direction (Kline, 2005), achieved by manipulating one variable with before-and-after measures of the criterion, and a supporting evidence base (Blunch, 2008). The study of Randolph, Arnett & Freske, (2004) suggested that dysfunctional attitude mediated the effect of depression and executive function on metamemory, measured by questionnaire. This was a non-experimental study, which used so-called 'causal' or path modelling as its analytic approach. In fact, this approach, which is based on covariance structure analysis, tests *a priori* relationships between variables, specified as a model, with directional 'causal' paths preassigned (Blunch, 2008:16). The supporting theory base for such models is seen to contribute to a sense of causal inference, supported by correlations and overall statistical fit of a model with the data (Schumacker & Lomax, 2004; Kline, 2005). If pre-existing theory is used to develop an *a priori* model, then causal assertions might be made with more justification, than if the model was derived *post hoc*, from the data (MacKinnon, 2008). MacKinnon defers to Sewell Wright, who developed the method of path analysis: *'The combination of knowledge of correlations with the knowledge of causal relations, to obtain certain results, is a different thing from the deduction of causal relations from correlations.'* (Wright, 1923:241).

In terms of understanding the factors that contribute to accurate self-report (by questionnaire) of memory, both studies have advantages. Julian, Merluzzi & Mohr (2007) provide a more definitive account of the causal importance of mood in self-assessments in cognition. Randolph, Arnett and Freske (2004) propose a useful way of thinking about the relationship of potentially causal variables, as well as some support for a relationship between self-report, low mood, negative attitude, and executive functions. Their model is presented in figure 1.4, and perhaps requires some interpretation.

The model is made up of observed variables (items in rectangles) and latent variables (items in oval enclosures). Latent variables are variables created from multiple observed measures, here the Tower of Hanoi and the Letter Number Sequencing tasks. These are proposed to represent a latent Executive variable, and directional arrows from the latent executive variable to each indicate that they load to a reliable extent on the factor. Measurement error for each variable is formally modelled so as to better represent the true relationships between variables; the *epsilon* notations indicate this. Straight arrows indicate parameter estimates, here indicating significant causal paths between variables. Finally, executive function and depression are correlated to allow for relationships between them that are unexplained in the model; these might include relationships related to MS pathology, personality factors or the effects of medications on each.

Despite the proposal in the model that Depressive Attitude mediated the Depression/Executive function relationship with metamemory (memory self-report) a number of features of this study are unclear.

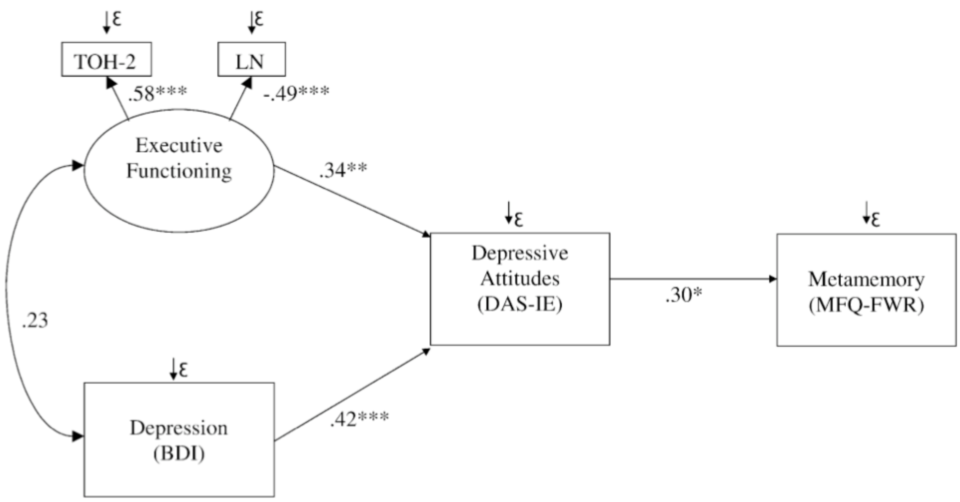


Fig 1.4. The structural model, of factors contributing to metamemory in MS, presented by Randolph Arnett & Freske (2004). The authors propose that both Depression (measured by the Beck Depression Inventory) and Executive Function (measured by the Letter Number Sequencing task and the Tower of Hanoi task) are mediated in their impact on Memory Function Questionnaire responses by negative self-oriented attitudes. No test of mediation is reported in the paper.

Aside from the model presented above, the relationship between self-reported and tested cognitive ability, other than memory, remains unclear. Beatty & Monson's (1991) memory questionnaire results suggest memory self-report is inconsistently related to cognitive abilities in MS, making this assessment approach unreliable. Randolph, Arnett & Freske (2004) propose that scoring on a memory questionnaire is associated with dysfunctional attitude, as a mediator of the effect of executive function and depressed mood. What is unclear is the theoretical basis of an executive function - self-report memory link. One possibility is a relationship wherein executive contributions to impaired memory performance are experienced, and reported, as a memory problem (Randolph, Arnett & Higginson, 2001).

Alternatively, understandings might be that in some way the executive role of *'sculpting of the response space'* (Fletcher, Shallice & Dolan, 2000: 404) is biased by negative attitudes. This could occur because of the interaction of cognition and emotion associated with the role of executive function and because reduced processing capacities, associated with depression, both constrain and bias evaluative processes (Bunce, Handley & Gaines, 2008; Baddeley & Hitch, 2007). Prefrontal functions may be generally involved in reflective abilities, including '*detailed, deliberative analysis, and maintenance of information while it is being evaluated'* (Nolde, Johnson & Raye, 1998). Taken together, there is support for both an affective and resource explanation that could characterise the impact of mood disorder.

One additional contribution to the self-report debate was a study by Randolph, Arnett & Higginson (2001) who investigated whether a recall task, an executive function measure and information processing abilities might relate to self-report of memory on a questionnaire in MS. The findings, in conflict with another study (Beatty & Monson, 1991) but supported by a later one (Randolph, Arnett & Freske, 2004), suggested that executive function and information processing speed, but not memory ability, are associated with memory questionnaire items relating to current memory performance in MS. This may support the contention that the memory deficits in MS are in acquisition; both in terms of speed and strategy deployment at encoding (Demaree et al., 1999; Chiaravalloti et al., 2003) and in this sense memory questionnaire reports may in part reflect these processes, rather than indices of forgetting.

1.10.2. Task-based Metamemory Judgment in Multiple Sclerosis

Cognitive factors appear contributory in task-based ('on-line') metamemory measures (Beatty & Monson, 1991; Scarrabelotti & Carroll, 1999), most notable in more demanding new-learning or episodic tasks (Pannu & Kaszniak, 2005). In the study of Beatty & Monson (1991) analysis of subgroups within the sample by levels of memory and executive deficit suggested that a *no memory impairment, no executive impairment* group were not different to controls on an episodic Feeling of Knowing task; all other subgroups - *no memory impairment, executive impairment; memory impairment, no executive impairment* and *memory impairment, executive impairment* were not reliably different from zero in their performance. One additional finding was that in a semantic task (the Nelson & Narens (1980) General Information Test), all groups performed as well as controls on the Feeling of Knowing measure, despite differences in recall performance for the task itself. The overall finding suggest that relevant cognitive domains were memory and executive function, and is supported by Pannu & Kaszniak's (2005) review of neurological populations' metamemory performance. Performance in episodic tasks do not compare with semantic tasks, the latter being potentially less useful in detecting difficulties with memory monitoring.

Heavily emphasised by Scarrabelotti & Carroll, is that impairments in metamemory are associated with decreased 'conscious processing' (Scarrabelotti & Carroll, 1999:1347); this conscious or intentional processing may support explicit memory interrogation and inferential reasoning. Supporting this suggestion was the study of Beatty & Monson (1991), who initially demonstrated that impairments in metamemory, compared to controls, were increasingly seen in those MS participants who had the dual memory/executive deficit profile, compared to those with just a memory deficit, or neither. Importantly for both studies, while there is an acceptance of the requirement for effortful, controlled processing, there is no focus on what might support, or constrain, the enactment of such processing. In the MS literature there is significant support for the idea that decrements in information processing abilities may underlie a range of complex cognitive tasks (DeSonneville et al., 2002; Denney et al., 2004; O'Brien et al., 2008; MacNiven et al., 2008), and the deficit is common (up to 60% of people with MS: Rao et al., 1993; Brassington & Marsh, 1998).

In the same way that limitations of processing ability (e.g. capacity, speed) might reduce abilities to effectively encode information or develop and deploy memory strategies (Salthouse, 1996; Demaree et al., 1999 Chiaravalloti et al., 2003; Bunce, 2003; Bunce & Macready, 2005; Lengenfelder et al., 2006) they might too be associated with difficulties with making accurate metamemorial judgments. Such judgments are proposed to be based on the maintenance of a dynamic model of proceeding cognitive operations, the organism's goal state, and to make comparisons necessary for ongoing discrepancy-reduction during learning (Van Overschelde, 2008).

Many metamemory judgments are, in explicit contexts (episodic research tasks, rehabilitation programmes) not automatised. Therefore it might be expected monitoring processes, being instituted as part of task-based processing, might require the recruitment of more effortful, non-automatic processing as task difficulty increases. In doing so, they may demand increasingly effective information processing (Shallice, Burgess & Robertson 1996; Schneider 1998; Scarrabelotti & Carroll, 1999; Garavan et al., 2002; Schneider & Lockl, 2008). The proposal for this study is that this might be evident as a positive relationship between information processing and the accuracy of metamemory judgment. Impairments in memory and executive function may lead to distortion in metamemory, but information processing limitations, common in MS, might underlie that distortion.

1.10.3. Mood and Task-based Metamemory in MS

Finally, the impact of negative attitudes, or depression, has not been investigated in task-based metamemory measures in MS samples. In one study in Parkinson's disease, it was suggested not to be relevant in the Feeling of Knowing performance of participants on a semantic memory (general knowledge) task (Coutler, 1989). Here, the proposal is that on an episodic task, which appears to be generally more challenging, the impact of low mood on effortful processing may disclose a relationship (Arnett et al., 1999a). Given that some metamemory judgments (e.g. Feeling of Knowing) are largely inferential, and given the results suggesting an interaction of mood and executive abilities in self-report, it may be that in conditions of difficulty, additional data for inferential judgments might include more global biasing factors, such as mood. Those conditions conferring complexity could relate to the nature of the judgment task or to impairments in cognitive domains typically considered relevant to the judgment.

The nature of this potential relationship might relate to the findings that depression can constrain working memory abilities in MS, and so might be relevant in demanding cognitive tasks (Arnett et al., 1999a, 1999b Arnett, Higginson & Randolph, 2001). This relationship has been proposed in older depressed people as a resource competition between attention-to-task versus attention-to-intrusive thoughts (Nebes et al., 2000; Christopher & MacDonald, 2005; Bunce, Handley & Gaines, 2008). This relationship has also been proposed in other neurological disorders, such as Huntington's disease (Nehl et al., 2001). Further, distortion may be added by the negatively biased appraisals associated with depression (Cavanaugh & Murphy 1986; Cavanaugh, Poon 1989; Lovera et. al., 2006).

Beatty & Monson, (1991) and Randolph, Arnett & Freske (2004) have similar goals in their studies - to evaluate factors that are relevant to metamemory in MS. Beatty & Monson (1991), address many aspects of metamemory; Feeling of Knowing for episodic and semantic memory tasks, a list learning prediction task and a memory questionnaire. The study did not include mood disorder as a variable to be considered, nor did it include a specific index of information processing abilities. A test of ‘verbal fluency’ used (the FAS test) was not included in the analysis, even though all subgroups differed from controls in the task, which might support its sensitivity in the sample, to pervasive information processing deficits in MS. Randolph, Arnett & Freske (2004) focused on memory-efficacy measures in their study, rather than task-based measures of metamemory. Important differences in the samples used are also apparent.

1.10.4. Samples

In terms of cognitive profile, Randolph, Arnett & Freske’s (2004) 48 participants presented with a low level of cognitive impairment - Letter Number Sequencing, a working memory measure according to Wechsler (1997), indicated 4% in borderline impaired range, and the Selective Reminding Test, a long term memory measure indicated 8% borderline impaired, - an unknown executive deficit (because the measure used is unstandardised) and 31% having a Beck Depression Inventory score in the mild to moderate depressed range (point prevalence of depression is cited as between 15% and 51% in MS, according to Mohr et al., 1997).

The memory performance, allowing for differences in testing methods, is considerably below the expectations from a review of 36 studies by Thornton & Raz, (1997), in two respects. There is a low level of working memory deficit; unusual because of its relationship with information processing and complex attention components, both common cognitive deficits in MS. Second, the long-term memory deficit is also sparse. Rao (2004) suggests 40%-60% of people with MS will display memory deficits. One issue is that the selective reminding procedure used by Randolph, Arnett & Freske (2004) may control for the impact of initial stimulus acquisition, therefore being a measure of delayed recall. As discussed when reviewing memory performance in MS, delayed recall is often not considered to be the key memory limitation in MS.

In Beatty & Monson's 1991 study of 45 MS subjects, 60% had a cognitive deficit of some kind based on performance below 5th percentile on the SECIMS (Screening Examination for Cognitive Impairment in MS; Beatty & Goodkin, 1990). No measure of mood disorder is reported in this sample.

1.10.5. Analytic Differences

This difference in approach relates to the aims of the studies. Randolph, Arnett and Freske (2004) aimed to model the performance of the sample as a group, whereas Beatty & Monson (1991) investigated subgroups of MS participants based not on disease subtype, but on cognitive performance, so as to clarify the memory-related and executive-related contributions to metamemory. This latter approach was proposed as useful when examining MS participants, because a single group based approach may have masked wide variance in performance (heterogeneity), which might be useful to interpret how some factors impact on metamemory, as opposed to others (Beatty & Monson, 1991).

Taken together, the approach of Randolph, Arnett and Freske (2004) might usefully model and characterise relationships between cognitive and mood factors on metamemory performance, but Beatty & Monson (1991) perhaps makes explicit the contributions, and the relative severity of cognitive performance required, to lead to metamemory impairments; in doing so there is additional focus on the quality of the relationships between, for example executive function and metamemory. However, the key findings, reported by Beatty & Monson suggest that the main effects of cognitive function group on episodic metamemory measures, were between controls and all other groups, rather than there being discrete homogeneity in each subgroup. It would seem that cognitive sub grouping might not have given the useful distinctions aimed for in terms of levels of severity relating to reliably different metamemory performance. A limitation in respect of the analyses is the number of individual analyses carried out, both planned and post hoc. In this regard, it is an exploratory study and the findings offer some direction for development on a more a priori basis, especially in how the field of research has advance since the study was published.

The structural equation modelling (SEM) approach, as used by Randolph Arnett & Freske (2004) allows for the development and testing of models of how variables are proposed to interact, and has some leaning towards causal interpretation; such models reflect causal assumptions on behalf on the investigator, and allow causal interpretations to be made on that assumption (Russo, 2008). An important aspect of this approach is that the to-be-tested models are theory driven (Russo, 2008). The SEM approach also allows for multiple measures being used to indicate a single construct (called a *latent* variable). Latent variables offer the benefit of a level of abstraction, by using the performance on a number of indicators to create the variable, for example in developing the variable *Executive Function*, Randolph, Arnett & Freske (2004) used two measures proposed to be indicative of the executive abilities.

Assumptions may have been made, based on theory, that the two measures were valid in representing this construct. These assumptions were not explicitly tested in the study; testing what is called the *measurement model* by means of confirmatory factor analysis typically achieves this. The purpose of this step is to test the validity of the proposed latent variables, especially relevant where unstandardised measures might be used, as in that study. SEModels with very small samples are also considered to have limited application (Schumacker & Lomax, 2004; Byrne, 2001; MacCallum, Browne & Sugawara, 1996). In Randolph, Arnett and Freske (2004) the sample was of 48 people with MS. Sample size considerations relate to the complexity of model, the more complex a model, the more parameters having to estimated with proportionally smaller amounts of data. Additionally, small sample sizes increase the chances of models being accepted (Schumacker & Lomax, 2004). Despite this, the selection of indices of statistical fit reported in the study might be considered unusual. Some have argued that the primary index of fit, the chi-square test, is a mandatory index (Barrett, 2007; McIntosh, 2007), and is not reported in this study.

Two final limitations are a failure to report a statistical test of mediation, such as the Sobel *z* test (Sobel, 1988; Preacher & Leonardelli, 2006). Although the reported fit statistics suggest an acceptable model, for mediation to be suggested it should be tested statistically. As mentioned, no confirmatory tests of the relationship between the indicators for executive function were carried out; instead as the two items were those that presented bivariate association they were proposed to reflect the same construct. Based on the earlier discussion, these could equally reflect processing abilities or working memory, not executive ability. The model does provide a useful starting point to extend the findings of the study and given issues with interpretation and method, could benefit from replication in this study.

1.11. Conclusions & Study Objectives

Cognitive Impairment is a common occurrence in Multiple Sclerosis with memory being one of the most commonly complained of deficits (Rao et al., 1993). Subjective reports of memory impairment have been investigated to only a limited extent and, in agreement with more general literature on memory self-report, have been found not to relate to tested memory abilities. Such subjective-report approaches to testing accuracy of memory appraisal are infrequently used in the contemporary research base on metamemory. However, they are frequently used in clinical practice, and judgment appears to associate with mood, rather than memory ability. Alternative approaches to assessing accuracy in memory monitoring include a range of judgments made concurrent with memory task performance. Both mnemonic and executive processes contribute to accuracy in this on-line monitoring of memory, using a range of cues to make judgments about past or future performance.

The impact of MS frequently leads to problems in each of the domains implicated in accuracy of metamemory judgments; mood, memory and executive function. Studies of metamemory in neuropsychologically impaired samples is limited; in MS only a few studies have addressed the issue of what makes for accurate metamemory judgments, broadly finding that questionnaire report frequently relates to mood, and metamemory monitoring, carried out during tasks, relates to mnemonic and executive abilities.

Limitations to drawing together the findings of these studies relate to differences between samples, measures and analytic approaches. None has sought to fully relate metamemory performance to current proposals about experiential, inferential or affective mechanisms involved in models of metamemory. Additionally, the impact of affective disturbance and decrements in information processing has not been assessed in the potentially demanding cognitive processing involved in metamemory monitoring in new learning.

Given the evidence that information processing capacities, notably speed, appears to compromise object-level cognitive performance it is relevant to considering its impact in meta-level performance also. In the two studies highlighted, samples have been relatively small (less than 50 participants) and both based in the US where social, cultural and medical care may impact findings. Based on these considerations, the potential for improvements in statistical methods, and the specific recommendations for future research into metamemory in neurological populations given by Pannu & Kaszniak (2005), the focus of this study will be to extend knowledge of metamemory in Multiple Sclerosis.

1.11.1. Aims of the Study

The general aim of the study is therefore to present a profile of performance on a range of metamemory measures of a UK community-dwelling sample of people with Multiple Sclerosis, the first time such a study has been carried out. This will focus on both self-report and a number of task-based monitoring judgments. Accuracy of judgments will be considered against mood disorder and implicated cognitive domains - memory, executive function and information processing ability. Analysis, using structural equation modelling, will consider how contemporary models of metamemory monitoring fit with performance of the sample and this will include replicating and extending the work of both Randolph, Arnett & Freske (2004) and Beatty & Monson (1991). The study will address the recommendations of Pannu & Kaszniak (2005): to relate different aspects of metamemory to executive function and a range of neuropsychological tasks, to further investigate the relationship between deficits in awareness of memory problem and depression, and to investigate of the relationship of metamemory inferred from questionnaire, and performance on experimental tasks. A final development of the research base will be the assessment of mediation of mnemonic or executive processes by information processing ability in metamemory accuracy.

Using structural equation modelling methods, a number of *a priori* models of the factors that contribute to metamemory accuracy will be investigated. For these models, a set of latent variables representing Memory, Executive Function, Mood and Information Processing will be proposed and tested for factor structure, using confirmatory factor analysis. The study will use four metamemory indices to evaluate metamemory accuracy; Memory efficacy (self-report), Judgment of Learning, Retrospective Confidence Judgment and Feeling of Knowing judgment. Each judgment will be made on episodic rather than semantic memory tasks, as new learning is one of the contextual factors for the study. Models will contribute towards understanding the factors that underpin memory-monitoring abilities in MS and provide some direction for both future study and intervention. Each of the *a priori* models is developed based on the review of the literature already presented, and models indicate what results are expected or unknown from that literature. The models presented are, at this stage, quite general, and will be specified to a higher level over the coming chapters as methods are developed. Development of each model will include specification based on the measures of accuracy selected, which is considered in *Chapter 2: Development of Methods*.

1.11.2. Memory efficacy models

The specific aim is to validate the findings of Randolph, Arnett & Freske (2004), that self-report memory is contributed to by executive function and mood, and that mood disorder mediates the executive contribution. Additionally, confirmation will be sought that memory self-report is unrelated to tested memory ability. The model will be tested using a more contemporary measure of mood disorder, and more psychometrically robust measure of memory self-report than reported by Randolph, Arnett & Freske (2004).

There are some additional differences between the two models; as the revised Beck Depression Inventory incorporates some of the negative attitudinal components associated with more cognitive characterisation of depression, a separate affective and attitudinal variable is not used in this model. Figure 1.5 summarises the *a priori* model; there are hypotheses of a positive correlation between Memory and Executive ability, a negative and statistically reliable parameter estimate between Mood and Metamemory indicating more depression being associated with reduced efficacy and a non-significant relationship between tested memory performance and metamemory. A mediational relationship will confirm if mood mediates the contributions of executive ability to metamemory.

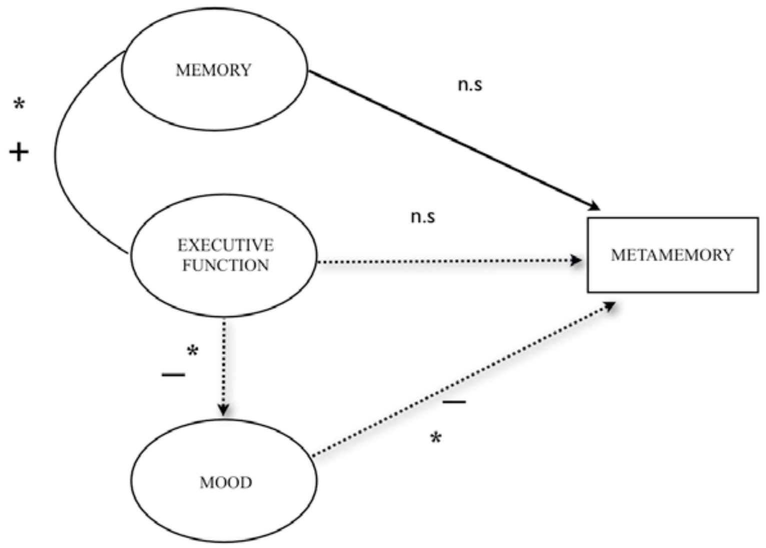


Figure 1.5. The a priori model of factors contributing to Metamemory (Memory-efficacy). Arrowed lines are proposed causal paths, with the dotted line indicating the mediational relationships to be tested. Curved lines indicate correlations. Parameter sign indicate the direction of expected relationships and * indicates those relationship expected to be statistically reliable at $p < 0.05$; n.s. = non significant estimates expected. Variables in oval enclosures are latent variables, measures for which will be confirmed in later analyses.

1.11.3. Judgment of Learning Model

The aim is to test the hypothesis that delayed Judgment of Learning accuracy is based on memory experience, and not associated with executive function. The model proposes that those who have better object-level memory ability are likely to benefit from the mnemonic factors thought to contribute to more accurate monitoring at delay on the Judgment of Learning - no dual-memory monitoring interference, the benefit of retrieval practice over multiple trials, and the benefits of covert retrieval attempts in making the judgment.

The model therefore proposes a positive and reliable association between accuracy in the delayed JoL and memory ability. This model will also extend the findings of Randolph Arnett & Freske (2004) to investigate if mood remains a contributory factor to metamemory in this task-based, rather than self-report measure of metamemory. Expectations for the model are that Executive Function will not contribute to accuracy to a reliable extent, and a contribution of Mood will be investigated. The model proposes positive correlations between cognitive items, but negative correlations between mood and cognitive items indicating more depression being associated with lower cognitive function (Landro et al., 2003; Haase et al., 2004). The a priori model is shown in figure 1.6

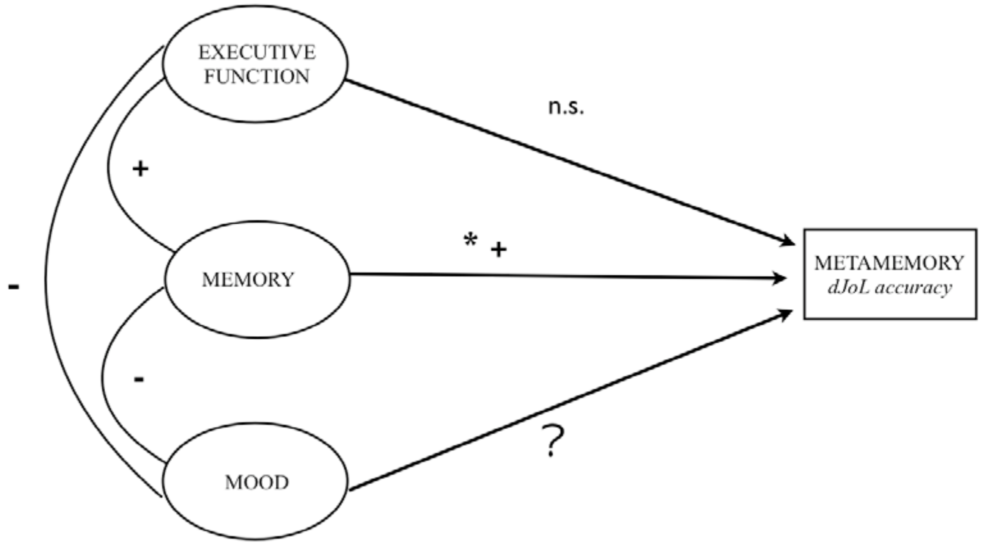


Figure 1.6. The a priori model of factors contributing to Metamemory (*delayed* Judgment of Learning: *d*JoL). Arrowed lines are proposed causal paths; curved lines indicate correlations. Parameter sign indicate the direction of expected relationships and * indicates proposed causal relationships expected to be statistically reliable at p <0.05; n.s. = non significant estimates expected. ? = parameter contribution to be investigated. Variables in oval enclosures are latent variables, measures for which will be confirmed in later analyses.

1.11.4. Retrospective Confidence model

The aim in this model is to test the hypothesis that Retrospective Confidence Judgment is based on memory experience and is not associated with executive/inferential processes. This is based on the proposal that this judgment, made on the product of recall offers many memory experience-driven cues towards accuracy; that privileged access informs the judgment and more robust Memory ability provides more such cues (Nelson et al., 1986; Koriat et al., 2008). Based on the discussed findings of Fu et al., (2005) Mood is not expected to contribute to accuracy to a reliable extent. Correlations between latent variables are expected to be as before; cognitive items correlating positively, and mood and cognitive items negatively. Figure1.7 summarises the expected relationships.

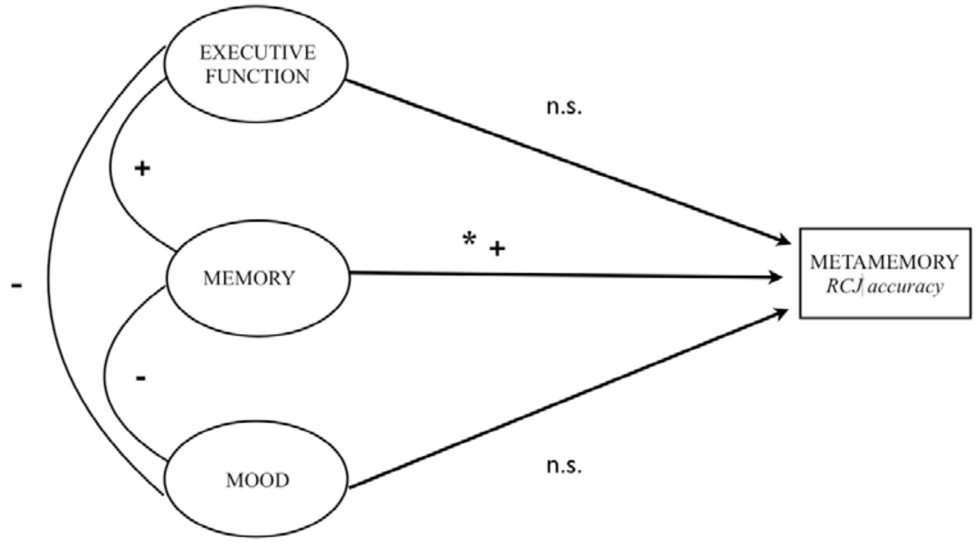


Figure 1.7. The a priori model of factors contributing to Metamemory (Retrospective Confidence Judgment: *RCJ*). Arrowed lines are proposed causal paths; curved lines indicate correlations. Parameter sign indicate the direction of expected relationships and * indicates proposed causal relationships expected to be statistically reliable at $p < 0.05$; n.s. = non significant estimates expected. Variables in oval enclosures are latent variables, measures for which will be confirmed in later analyses.

1.11.5. Feeling of Knowing model

The primary aim is to retest the findings of Beatty & Monson (1991) that Memory and Executive Function are reliable contributors to Feeling of Knowing judgment accuracy. Findings in other neurological samples appear to suggest that Executive ability is the primary contributor in Feeling of Knowing accuracy (Shimamura & Squire 1986; Pannu & Kaszniak, 2005). Assessing the relationship between mood and accuracy will extend those findings. The model is shown in Figure 1.8.

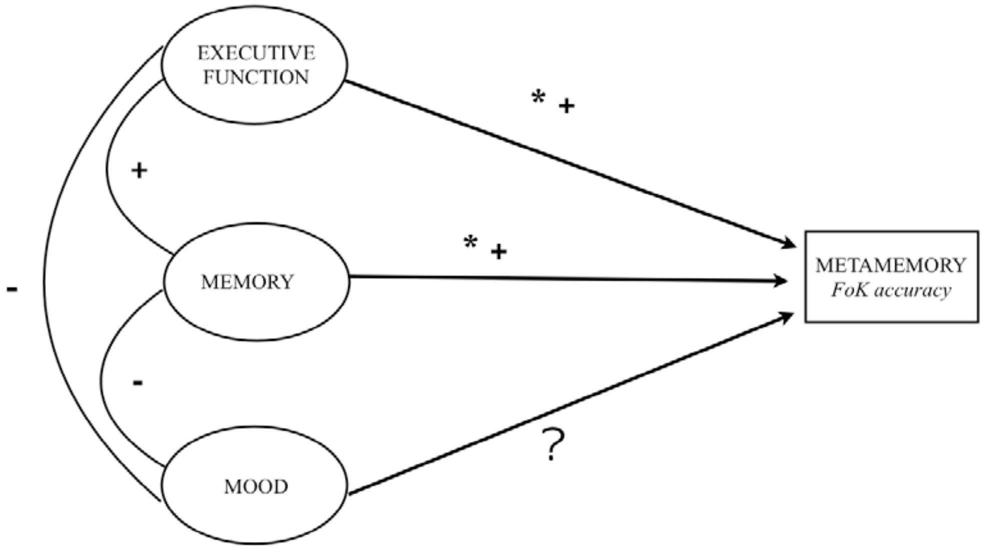


Figure 1.8. The a priori model of factors contributing to accuracy in Metamemory (Feeling of Knowing: *FoK*). Arrowed lines are proposed causal paths; curved lines indicate correlations. Parameter sign indicate the direction of expected relationships and * indicates those relationship expected to be statistically reliable at $p < 0.05$; ? = parameter contribution to be investigated. Variables in oval enclosures are latent variables, measures for which will be confirmed in later analyses.

1.11.6. Information Processing mediation models

A final set of models will be tested for each of the three task-based judgments to investigate a mediational role for information processing in accuracy. These models will tests for mediation of Executive Function by Information Processing and assume that under easier judgment conditions, i.e. those with available memory experience, there is low need for controlled processing resources, so that Memory is not mediated by Information Processing. In the low availability Feeling of Knowing task, Information Processing may mediate both Memory and Executive Function, to the extent that it facilitates the process of monitoring amongst the more distal cues that are used in making accurate judgment. The generalised information processing as mediator model is shown in figure 1.9.

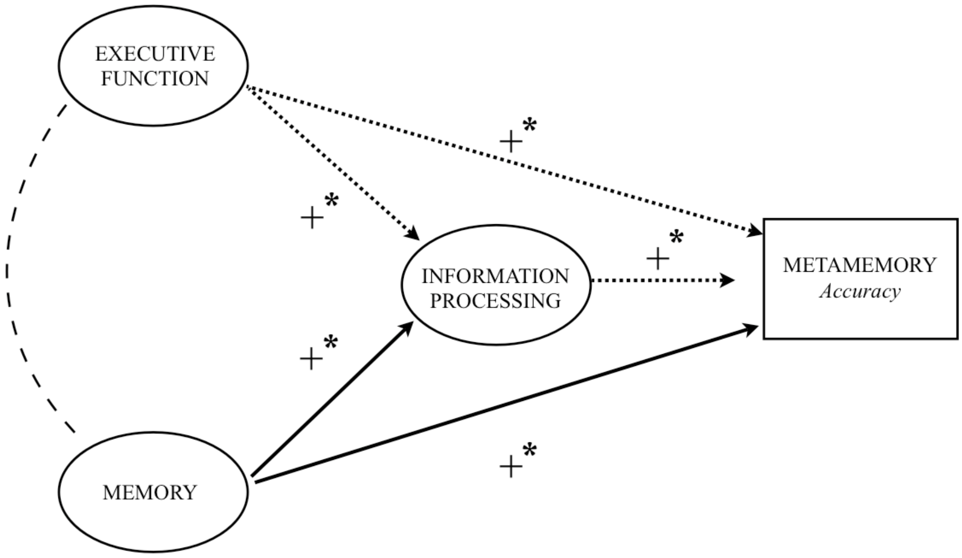


Figure 1.9. The generalised *a priori* model of Information Processing meditating factors contributing to accuracy in Metamemory. Arrowed lines are proposed causal paths; the dotted causal paths indicate the mediation relationship of interest for Executive Function. Curved lines indicate correlations, here this may not be applied if both Memory and Executive Function mediation can be established, as their shared variance in information processing is modelled thorough the mediational relationship. Parameter sign indicate the direction of expected relationships and * indicates those relationship expected to be statistically reliable at p <0.05;Variables in oval enclosures are latent variables, measures for which will be confirmed in later analyses. For the purposes of generality, Mood is not included in this model.

Chapter 2: Development of Methods

2.1 Introduction

This chapter will focus on issues relating to measurement for each of the variables in the study. Each selected test will create an *observed variable* (Schumacker & Lomax, 2004), that is, each acts as an index of a proposed domain, such as *Memory*. These individual variables are useful for the description of sample performance, most of them being standardised measures, for which norms are available in the literature.

The main purpose of the study is to examine the relationships between the domains that these measures index; *Memory, Mood, Executive Function, Information processing* and *Metamemory*. Structural Equation Modelling (SEM) is pursued as an appropriate method to achieve this because it allows for the use of multiple indicators of each domain, and the testing of interactions between these (Kline, 2005; Weston & Gore, 2006). The first step in modelling these interactions is to create descriptors of each domain, termed *latent variables* (Schumacker & Lomax, 2004). Latent variables are created from combining observed variables that are proposed to load on the same construct (Schumacker & Lomax, 2004).

In selecting observed measures to construct a latent variable, testing that they do satisfy as indicators of a latent construct is also planned as part of the statistical methods. In the later analysis of structural models, the relationships between these latent variables are then examined. The initial step in constructing latent variables is to review available measures and consider what published evidence is available for their reliability and validity as measures of the domains of interest in the sample under study.

2.2. Selection of Latent Variable Measures.

A general issue which pervades test selection for this study is to develop a range of tests that can be completed by a heterogeneous group of people with Multiple Sclerosis (MS), taking into account the typical range of problems associated with the disease – visual disturbance, motor disorders (including speech and upper limb problems) and fatigue (Lezak, Howieson & Loring, 2004).

The aim was to avoid giving tests ‘on which failure is both inevitable and uninterpretable’ (Lezak, Howieson & Loring, 2004 pg 250). Therefore, visual complexity, motor demands and test length were considerations alongside the psychometric properties of each test to ensure appropriate measures were selected to gather the planned data. Pragmatic concerns included qualifications of the researcher, resource issues, relevance to clinical practice, and given that participants might need to be assessed in their own home environment, test portability (Auger, Demers & Swaine, 2006).

An important additional issue was the level of *practice effect* with selected tests (Beatty, 1999). Practice effects occur where experience with doing a test in the past supports better performance in the future, because of that previous experience (and not because of change in abilities; Field, 2005 pg 273). While this is not a longitudinal or repeated measures study, the fact that participants may have self-selected for involvement in the study, on the basis of some orientation towards their own cognitive function, may increase the likelihood of having been assessed in the past. Many of the cognitive assessments used in MS are routine because of the existence of guidelines for cognitive screening and commonly used screening batteries (Peyser et al., 1990; Rao, 1990). Therefore, consideration was also given to the availability of alternate form versions and the evidence about the length of practice effects (e.g. Wisconsin Card Sorting Test for Executive Function; (Beatty, 1999; Barker-Collo, 2005). The process and outcome of measure selection will now be described.

2.2.1 Test Selection for Memory.

As discussed in the previous chapter, there are a number of aspects of memory abilities that have been shown to be impaired in MS (Rao, 1995; Landro, Sletvold & Celius, 2000; Mohr & Cox, 2001). Testing memory performance requires that a number of manipulations be considered. Broadly, these relate to the length of delay between presentation and test of recall, immediate or delayed, and the particular retrieval demands; free recall, cued recall or recognition (Lezak, Howieson & Loring, 2004). One complication with a study of judgements about memory is that it requires tasks, about which participants can also make judgements about their memory performance. Therefore the selection of memory tasks is required to meet a number of criteria; have variable length of delay (immediate recall, delayed recall), have variable retrieval demands (recognition, free and cued recall), be auditory/verbal or of reduced visual demand, have acceptable reliability and validity and have some normed or reference group with whom to compare performance. The proposed *Memory* latent variable will here represent delayed recall ability.

One issue with tests of delayed recall for people MS, is that they may be confounded with the impact of the speed at which to-be-remembered information is presented (Demaree et al, 1999; Arnett, 2004). This typically has an impact on initial acquisition (Salthouse, 1991), and can appear to be a recall problem, when in fact the information was never acquired in the first instance. Speed of information processing is a problem for people with MS (Diamond, et al., 1997; DeLuca, et al., 2004) and this may be independent of memory performance in that when controlled for, memory performance can be normal (Demaree, et al., 1999). It is therefore of interest to measure delayed recall that is not entirely dependent on processing speed. Additionally, processing limitations in general can be confounded with recall in tests that offer a single presentation of large amounts of to-be-remembered information (Lezak, Howieson & Loring, 2004), regardless of the speed of that presentation. Capacity limits impact on how much information can be acquired with a single presentation even in non-neurologically impaired adults (Baddeley & Hitch, 1974; Baddeley & Hitch, 2001). A balance has therefore to be sought in not biasing the measures which define the latent variable *Memory* by entirely removing the impact of information processing - it is after all, an aim of the study to investigate the mediating role of information processing on the relationship between memory and metamemory. At the same time, concerns about adequate discriminant validity needs to be considered in not selecting a memory measure that, in reality, *only* reflects information processing abilities. One benefit in proposing a latent variable for this domain is that more than one measure can be used to capture performance.

A final feature of an appropriate memory measure is that it requires scoring that is devoid of gist-based recall, so that predicted versus actual recall can be reliably compared; so performance can easily be deemed correct or incorrect. For example, if a stimulus story had information about a *milkman knocking on the door* and a participant reported a *postman knocking on the door* it would introduce unwanted complexity to the scoring, because the decision about correctness becomes difficult - information was half-remembered, or the participant controlled the level of generality in their answer so as to maintain accuracy (Koriat, 2000). The exclusion of tests using this type of recall will be a trade-off in terms of external validity, to maintain internal validity. This excluded tasks like the Story Recall and Logical Memory Test from the Rivermead Behavioural Memory Test (Wilson, et al., 1999) and WAIS-III (Wechsler, 1997a) both offer scoring for gist-based recall (synonyms get points as well as word-for-word recall).

Consideration of the issues associated with a single presentation of stimulus items, gist based recall and slowed information processing, led to a review of a number of tests that offered multiple learning trials before imposing a delay prior to testing recall. Typically, these tasks are list-learning paradigms, which can be considered a sensitive measure of memory, because they do not offer the support of context or gist, which might aid in story-recall tasks (Lezak, Howieson & Loring, 2004). This category of memory tests includes tests such as the Selective Reminding Test (SRT: Buschke, 1973), The California Verbal Learning Test (CVLT; Delis, et al., 1987) and the Rey Auditory Verbal Learning Test (Rey, 1958 cited in Spreen & Strauss, 1998: 326; Lezak, Howieson & Loring, 2004).

The Selective Reminding Test gives participants a reminder of the words they have not recalled from the first and each subsequent trials of a 12-word list; therefore offering ‘selective reminding’. Multiple trials of a 12-word list are given and the selective reminding continues until full list recall is achieved on three consecutive trials. There is then a 30-minute delay before recall is tested. In all, the test takes about 30-minutes to complete (Spreen & Strauss, 1998). It has been used in, and has been recommended for, the assessment of memory in MS (Beatty, 1999; Beatty et al., 1996a; Beatty et al., 1996b). However, with a 30-minute administration time and a design offering multiple indices of verbal memory (e.g. long-term storage, long-term retrieval, consistent long-term retrieval) this task is more complex than necessary for this study’s purposes. Lezak, Howieson & Loring (2004) also point out that this test has not been compared empirically to other learning tests to establish if it offers a better assessment of memory ability.

The administration time of the California Verbal Learning Test (CVLT; Delis, et al., 1987), a similar list-learning task, runs to about 35 minutes (Spreen & Strauss, 1998), and scoring reflects the strategies related to remembering perhaps more than the recall performance itself. Given the range of assessments planned in a single testing session, it takes too long to administer, and is again more complex than necessary for the purpose of this research.

A shorter administration time and reduced complexity is offered by the third of the repeated presentation list-learning tasks - the Rey Auditory Verbal Learning Test (RAVLT or AVLT; Rey, 1958: cited in Spreen & Strauss, 1998:326; Lezak, Howieson & Loring, 2004, for the English version). The test has been used with an MS sample before (Bravin, et al., 2000), as an index of retrospective recall. Additionally, it does not suffer from SRT problems, such as being ‘often discouraging, certainly boring’ (Lezak, Howieson & Loring, 2004:443) or from performance being confounded because the stimulus items are already categorised, aiding recall for the learner, as in the CVLT (Lezak, Howieson & Loring, 2004). Extensive norms are available (Spreen & Strauss 1998), administration time is 10-15 minutes and alternate forms are available. It was therefore considered one appropriate test to use in this study, as the first of two tests of delayed recall.

The second recall test had two criteria to fulfil - it was a test of recall after a delay and it could provide data that would contribute toward the assessment of metamemory. Additionally it would have to be more demanding in terms of the learning support it required, to balance the repeated presentation of the AVLT. A Sentence Memory paradigm was investigated to fulfil these criteria, as it has a long history of use in metamemory experiments.

The Sentence Memory Task (Nelson & Narens, 1980; Nelson, 1984; Shimamura & Squire, 1986) is not a standardised test of memory function, instead it is a testing paradigm used for generating data on recall performance and on accuracy of metamemorial predictions. Versions of this paradigm have been used in many metamemory studies in the past, including in MS (Beatty & Monson, 1991; Scarrabelotti & Carroll, 1998; Scarrabelotti & Carroll, 1999).

Sentences for the test were taken from a set of incomplete sentences produced by Bloom & Fischler (1980) and the sentence completion words are based on those generated by a London-based sample of 73 non-neurologically impaired adults, using the Bloom & Fischer stimuli (Arcuri, et al., 2001). To avoid participants being correct by guessing the likely high frequency response during the test, the lower frequency generated completions were selected. The full set of test materials are presented in Appendix A.

The task structure of the Sentence Memory Task is based on a test paradigm used by Nelson (1984), Shimamura & Squire (1986), and where participants are given a number of sentences to read with their attention drawn to the final word of each sentence. These sentence completion words are learned from a single exposure and then, after a delay, participants are re-presented with the same sentences, with the final word of each sentence missing. A participant's first task is to provide the word they remember to be the word that completed each sentence. The performance on this recall portion of the task will be used as the second measure of delayed recall.

The test also meets the second criteria of providing data for metamemory judgments. Use of the test to derive measures of metamemory (as opposed to memory) is discussed later in this chapter. While the Sentence Memory Task paradigm is appropriate for the study, as a paradigm, it does not stipulate the exact materials (i.e. the sentences) to be used. A full outline of the test structure and materials used in this study is given in Chapter 4: *Methods*. Administration time for the task is about 30 minutes, including the task elements relating to metamemory. This was considered acceptable, given the amount of information the process yields.

In summary, two memory tasks were found to be appropriate to model the latent variable 'Memory', which will be based on 20-minute delayed recall performance in the Auditory Verbal Learning Task and 60-minute delayed recall on the Sentence Memory Task. Both tasks reflect the inclusion criteria for study tasks; the AVLT has equivalent formats for those who might have experienced it before, the Sentence Memory Task will be new to all, as it has been compiled for this study. Each is relatively quick to administer and yields interpretable data for the study. While only the AVLT has adequate norms, the Sentence Memory Task primary function is in providing the basis for assessment of metamemory, rather than memory. The appropriateness of the Sentence Memory Task as an additional assessment of memory will be explicitly tested during the confirmatory factor analytic testing of validity. Figure 2.1. summarises the proposed formation of the latent variable for memory; convention in confirmatory factor analysis and structural equation modelling indicates a latent variable by an oval enclosure, and an observed variable by a rectangular one (Schumacker & Lomax, 2004; Kline, 2005).

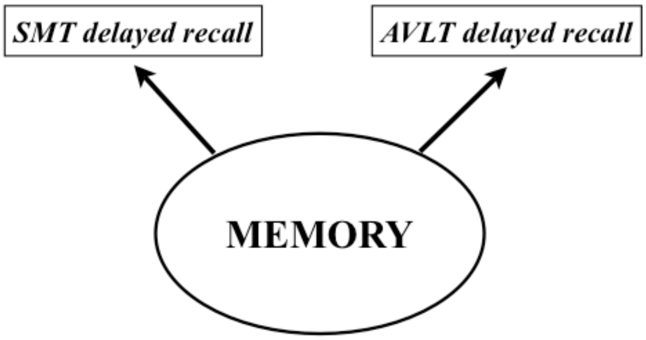


Figure 2.1 summary of observed measure for the proposed latent variable *Memory*. Latent variables are indicated by ovals and observed variables by rectangles.

2.2.2. Test Selection for Information Processing.

Given that one of the aims of the study is to investigate the impact of information processing abilities on metamemory, the collection of data reflecting this process (or processes) bears consideration – the main issues being the validity and reliability of measures used, as well as how best to operationalise the construct at the latent variable level.

A guiding principle in selection of measures for the information processing latent variable is the proposal that this latent variable represents a key process underlying, perhaps mediating, the functioning of memory and other abilities, in their relationship with metamemory. Here, the variable is proposed to represent speed of processing and capacity to maintain and manipulate information 'on-line' (Gontkovsky & Beatty, 2006). The latter some would term 'working-memory' (Baddeley, 1996; Andrade, 2001a). That these items have some unidimensionality, in the sample of people with MS under investigation, will be tested in the confirmatory analytic process.

Information processing abilities in MS, and their measurement, have been widely investigated (Archibald & Fisk, 2000; Fisk & Archibald, 2001; Rosti, et al., 2006; Williams et al., 2006; Gontkovsky & Beatty, 2006; Parmenter, et al., 2007). Two issues arise from the literature – a speed-accuracy trade off in measurement, and the expansion of the information-processing concept to include attention and working memory items (Demaree, et al., 1999; Gontkovsky & Beatty, 2006). Typically, the published literature focuses on two main indices of information processing in MS – the Paced Auditory Serial

Addition Test (PASAT; Gronwall & Sampson, 1974; Gronwall, 1977) and the Symbol Digit Modalities test (SDMT; Smith 1982); both were considered as potential measures for this study.

Some advantages of the PASAT include the reduced motor demands of the task, which involves the mental serial addition (of two numbers, the currently-presented plus the previously-presented; that is, n + n-1) at a paced speed (increasing from a digit every 2.4 seconds to every 1.2 seconds). The participant verbally reports the sum of the two numbers, while numbers continue to be delivered through headphones for the next serial addition. This test has been considered an 'aversive' test (Fos, et al., 2000), which some people refuse to do, cannot do, or stop doing because of the difficulty (Spreen & Strauss, 1998: 248; Lezak, Howieson & Loring, 2004; Gontkovsky & Beatty, 2006).

While the instructions are typically presented on audiotape, there can be a need to repeat these, or provide them in the form of an oral or written version too (Spreen & Strauss, 1998). Practice effects have been noted, both within (Barker-Collo, 2005; using an MS sample), and between, sessions (Gronwall, 1977). The test takes, assuming no repetition or demonstration of instructions is required, approximately 15-20 minutes to complete (Spreen & Strauss, 1998). Additional limitations include the potential for difficulties distinguishing between problems with information processing demands of the task (speed of responses) and dysarthria (Spreen & Strauss, 1998); another is the numerical ability of the study sample, as perceived or actual numerical ability can have an effect on participation. Given the potential for difficulties with the test, an alternative to this test, which is shorter and less complex to administer, but with proven value for use in MS, was therefore was sought.

Three shorter tests, also developed for the measure of information processing, are the Digit Symbol-Coding task (WAIS III, Wechsler, 1997a), the Coding subtests of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, et al., 1998) and the Symbol Digit Modalities Test (SDMT: Smith, 1982). Each test requires the matching of novel 'hieroglyphic-like' symbols (Wechsler, 1997b:2) to numbers, using a key that shows which numbers and symbols go together. The aim of the tasks is to correctly match as many as possible in a set time, e.g. 90 seconds for the SDMT. One limitation, for an MS sample, of both the Digit Symbol-Coding and the Coding subtest of the RBANS is that it requires the symbols to be written down as the response, offering the potential of compromised motor function as a confounding issue in the time-based score. The SDMT also requires a written response, but an oral administration, where the participant calls out the numbers that go with each symbol is also permitted. Having an oral-only response option, as well as a test that could be completed in approximately two minutes was considered worthy of pursuit, so the psychometric properties of the test were investigated.

The oral version of the SDMT (Smith 1982) has been investigated for use in MS as an index of processing abilities (Gontkovsky & Beatty, 2006) and is recommended for use in preference to the PASAT or the RBANS coding because of the oral response option. In terms of visual acuity, the stimuli are likely to be sufficiently large (Schear & Sato, 1989; Gontkovsky & Beatty, 2006). Recency effects appear not to affect performance (Lezak, Howieson & Loring, 2004) and the test, when used on people with MS who reported memory problems, correlated better with non-memory cognitive domains, rather than with memory performance (Randolph, Arnett & Higginson, 2001), suggesting it is not a memory related task, perhaps attesting to a relative purity as an information processing measure. Generally it has been proposed that the test is a sensitive measure of brain damage (Smith, 1982; Spreen & Strauss, 1998) and some propose it was the single best measure of information processing speed out of two alternatives (Ponsford & Kinsella, 1992; Drake et al., 2010).

Specificity for the instrument was reported as 0.82 by the author (Smith, 1982), and it has been demonstrated to correlate positively with neuropsychological performance, and negatively with mood disorder (Parmenter et al., 2007). A cut off of 55 correct responses demonstrated a sensitivity of 0.82 in an MS sample (Parmenter et al., 2007). The oral and written versions also correlated at r = 0.82 (Smith, 1982). The SDMT was therefore considered one appropriate test of information processing in people with MS for this study, but given the aim of reducing measurement error, and better characterising the latent variable, others were sought.

As discussed in Chapter 1, in selecting any tests of information processing ability, it must be recognised that definitions of information processing, vary considerably. Some describe the SDMT as a test of sustained attention (Boringa et al, 2001), or visual attention (Beatty, 1999), information processing speed (Gontkovsky & Beatty, 2006) or attention and scanning (Sheridan et al., 2006). These differences about definitions reflect an expansion of the information-processing concept to include attention and working memory items and not just speed (Demaree, et al., 1999; Bunce & Macready 2005; Gontkovsky & Beatty, 2006).

As part of developing this latent construct, working memory capacities will be measured; if the SDMT is considered a measure of speed of information processing, then it is proposed that measures of processing capacity and mental manipulative abilities will also be included so as to develop a latent variable representing a range of operations relating to processing ability. This is because of the aim and theoretical orientation of this study, that this latent information-processing variable underpins the performance of other, more cognitively intensive mental operations, in the processing that supports metamemory. The lack of agreement about what processing capacities might or should include also supports selecting potential measures and using the confirmatory factor analyses to clarify their relationships. Therefore, measures of processing span (capacity) as well as manipulative processing were also considered for inclusion

The ability to repeat progressively longer stimulus lists is considered a measure of immediate recall, or information processing capacity, both often considered a quantity dimension of attention (Spreen & Strauss 1998; Lezak, Howieson & Loring, 2004). Two standard measures were considered - the Digit Span Task (looking at how much information can be maintained for immediate recall) and Letter Number Sequencing task (looking at abilities to maintain and manipulate information for immediate (i.e. no-delay) recall).

Span tasks come in a number of forms but typically fit into three types – span, supraspan and what Lezak, Howieson & Loring term ‘mental tracking’ (2004:360) span tasks. Straightforward span tasks start at easier capacity limits, e.g. two numbers to be repeated back, increasing in length from there. Supraspan tasks start with more than is possible to retain e.g. 15 items, and are therefore perhaps confounded with memory processes from the outset (Lezak, Howieson & Loring, 2004; the first trial of the AVLT is an example of a supraspan task, with 15 items). The mental control tasks have requirements for maintenance and mental manipulation of the given information, but, like simple span, they work from below capacity limits, increasing by 1 stimulus on each trial. Thus, supraspan tasks, because of their mnemonic basis, will not be included as measures of processing abilities. Three simple-span tests are reviewed first: Visual Span, Sentence Span, Digits-Forward task (WAIS III; Wechsler, 1997b).

Visual span tasks, such as the Corsi Block test (Milner, 1971) were discounted as the requirement for copying the block tapping pattern of the examiner may instead be sensitive to difficulties with the reaching and tapping blocks; even if achievable, it might be considered an interference (i.e. be attention demanding) in a person with a movement disorder, which could attenuate performance. The motor demands of the task are therefore likely to be confounded with recall performance, making this task inappropriate for use with MS participants. Difficulties with oculomotor control, diplopia or nystagmus, common in MS (Reulen, Sanders & Hogenhuis, 1983; Starck et al., 1996), might also lead to difficulties with the visual pursuit required in such tasks.

Sentence repetition tasks require participants to repeat sentences that become increasingly longer (e.g. more syllables, more content). One noted confound is how meaningfulness, and familiarity, contribute to abilities at repetition, when compared to span for unrelated items (Lezak, Howieson & Loring, 2004). One other limitation is the importance of exact repetition; adjustment of pronouns is an example that Lezak, Howieson & Loring (2004) discuss, along with the impact of regional or ‘ethnic’ English on exact repetition. As there are other sentence-based tasks in the study this was considered inappropriate so as to avoid the potential for interference between tasks.

The Digit Span forwards task, with numerical stimuli, offers some benefits in terms of the limitations of both tests discussed above. In addition, being part of the WAIS it has had considerable investigation into its dimensionality and internal structure, mainly confirming that it described a simple attention capacity measure (Spreen & Strauss, 1998), part of ‘freedom from distractibility’ measure (Sherman et al., 1995) also considered ‘working memory’ (Wechsler 1997a). It was therefore selected as an appropriate information processing measure for this study, reflecting the capacity aspects of processing.

Of interest, the Letter Number Sequencing task (LNS; Wechsler, 1997b) also correlated highly with a working memory/freedom from distraction factor, along with the Digit span task (LNS = 0.85; Digits = 0.83; The Psychological Corporation 2002:78). The Letter

Number Sequencing (LNS) task was therefore investigated further as a third, and more complex measure of attention / processing ability relevant to the aims of this study. The LNS task (along with Digits and Arithmetic) forms part of the Working Memory Index of the WAIS-III, supported by factor analyses (WAIS technical manual: table 4.6; Saklofske, Hildebrand & Gorsuch, 2000; Arnau & Thompson 2000). The factor structure also appears stable across age ranges (Wechsler, 1997a).

Though the LNS concords with what Lezak, Howieson & Loring (2004:360) termed ‘mental tracking’, working memory, or attentionally demanding tests. Others have considered it part of an executive ability (Randolph, Arnett & Freske, 2004). The main reason for this is theoretical orientation. In line with the proposal in this study, it will be considered a measure of working memory or ‘on-line’ processing, but a conclusion about whether it loads more on (has a greater statistical relationship with) an executive versus information processing/mental abilities factor can only be drawn after the confirmatory factor analysis has been carried out. It fulfils the criteria as an appropriate measure because of its short administration time and verbal presentation, but is selected with cognisance that it is unclear whether it would load on an information-processing or executive function factor in this sample. The expectation, to be tested in confirmatory factor analysis, is that it will load with the other two selected measures, SDMT and Digit Span Forwards tasks, on an information-processing factor.

Alternative tests considered good indicators of the same functions include the PASAT (Gronwall & Sampson, 1974), the limitations of which were discussed earlier, and the Stroop Test (Stroop, 1935; Jensen & Rohwer, 1966). A limitation of the Stroop Test is the significant reliance on visual stimuli, which can consist of over one hundred colour names on a single A4 page (Sacks et al., 1991). Those with visual acuity problems may actually be at an advantage as the ability to easily read words increases the difficulty of ignoring the written word and attending to the colour of the ink in which the word is written (Lezak, Howieson & Loring, 2004). Conversely, the incidence of colour vision disturbance in people with MS (Ashworth, Aspinall & Mitchell, 1989) is likely to make colour naming more difficult.

In summary then, the latent variable of *Information Processing* will be constructed from the Symbol Digit Modalities Test, Digit Span Forwards and Letter Number Sequencing Test.

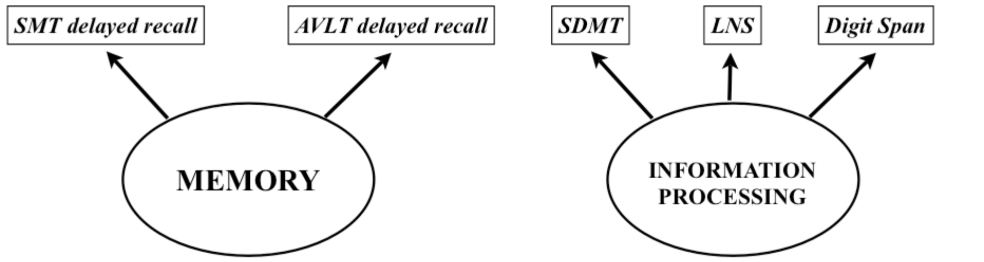


Figure 2.2. Summary of observed measures for the proposed latent variables of *Memory and Information Processing*

2.2.3. Test Selection for Executive Function.

One of the difficulties in testing executive functions is the diversity of component processes included in the construct (Spreen & Strauss, 1998; Miyake et al, 2000a). This leads to a number of issues from a measurement perspective. The first issue is what to measure in order to represent the construct. Broadly the skills involved include, at the behavioural level; planning, problem solving, regulation, monitoring and self-correction (Luria, 1980; Prigitano, 1999) The second issue is how reliably can these skills be measured; the multidimensionality is compounded by the need for low structure in the testing situation to allow for the demonstration of 'discretionary behaviours', that is executive abilities, on the part of the person being tested (Lezak, Howieson & Loring 2004:612).

Cognitive Neuropsychology has proposed a number of models of the executive system, one of which presents the idea of a Supervisory Attention System, which is engaged by novel (non-routine), complex or control-requiring tasks (Norman & Shallice, 1980; Shallice 1988; Miyake et al, 2000a). The focus on novelty, or *non-routineness,* is an important consideration because of how it relates to assessment procedures.

Miyake, et al., (2000a) and Royall et al., (2002) in their consideration of the key dimensions of executive control, posit *response inhibition, working memory, rule discovery, updating* and *set shifting* as some of the key areas in executive function. Because of the multiple components of executive function, difficulty with being sure what is actually being tested (Rabbitt, 1997; Burgess, 1997), and the nature of the tests themselves (low structure, requirement for novelty Burgess, 1997) have all been proposed as reason for low reliability and poor inter-test correlations (Burgess, 1997). As novelty is so important, test-retest reliability would be expected to be low (Beatty, 1999). Because of low structure, and therefore space for varieties of approaches to the task, internal validity may also be low (Rabbitt, 1997; Miyake, et al., 2000b).

Aside from assessment techniques needing to be novel and having low levels of structure, which might aid performance, a second relevant issue in selecting measures of executive function is whether the measures are diagnostic of, or just sensitive to, disorders of executive function. Because of the proposed multidimensionality of the construct, selection of tests will aim for indicative measures rather than a whole battery of tasks. In a clinical situation, a full battery may be more appropriate where diagnostic concerns are primary. In the research context, where fatigue and testing-session length are important considerations, indicative testing is proposed to be sufficient. Full batteries of tasks, such as the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1997; Norris & Tate, 2000) were not therefore pursued as appropriate measures for this study.

A final consideration relates to what Miyake et al., (2000b:52) term ‘task purity’. Studies have shown differences (and therefore possibly separable sub-functions) in performance between executive tasks (Miyake et al., 2000a). These could be accounted for by differences in the cognitive domains managed, or used, by the executive system – e.g. language abilities (Miyake et al., 2000b; Godefroy, 2003). The implications of this support the use of a number of measures to form the latent concept of executive function, as well as measures emphasising both *executive in verbal*, and *executive in non-verbal* domains (Hunt & Kingstone, 2004). As with other assessments, they also need to be relatively short and not heavily reliant on complex visual or motor mediated responses.

The Stroop task, a frequently used measure of executive function, has been shown to be sensitive to frontal lobe functioning if comprehension or visual search deficits are controlled for (Stuss et al., 2001). In one study (De Frais, Dixon & Strauss, 2006) it loaded, along with the Colour Trails task, on the same factor as the two other tests, the Hayling Sentence Completion test and the Brixton Spatial Anticipation test. The limitation of the Stroop’s presentation format was discussed earlier when considering this test as a potential for testing information processing abilities. However, many also consider this test as sensitive to executive function (De Frais, Dixon & Strauss, 2006) and as such the reverse confound is proposed. MacNiven, et al., (2008) warn the use of tests where reaction time is a primary measure may not be appropriate for people with MS, because of the possibility of confounding with information processing abilities. It is likely therefore that in a factor analytic study, Stroop performance could easily cross-load between executive and information processing factors, making the modelling of one factor mediating the other more complex than is necessary, when an alternative measure might avoid such cross-loading effects.

Because of its use in a key paper on metamemory in MS (Randolph, Arnett & Freske, 2004), and their popularity as tests of executive function (e.g. Shallice, 1988; Manchester, Priestley & Jackson, 2004), Tower tests (Towers of Hanoi, London and others) were also considered. Of note, Randolph, Arnett & Freske (2004) used a computerised version of the Tower of Hanoi (ToH) task, presumably to reduce the impact of motor function limitations, in their sample of people with MS. However, this led to not having norms for performance on the task for comparative purposes. Tower tests do have a good track record for use as indices of executive function (Welsh, Satterlee-Cartmell & Stine, 1999), more specifically of planning abilities (Arnett et al., 1997; Arnett, et al., 2001; Lezak, Howieson & Loring, 2004). Such tasks are typically scored on the number of moves it takes to move the pieces of the tower (balls, discs) from an initial position to a desired position, with different towers having different levels of complexity. At certain points during the task it is required that counterintuitive moves be made, for the sake of future success, some have suggested that inhibitory processes are important (Goel & Grafman, 1995; Welsh, Satterlee-Cartmell & Stine, 1999; Miyake et al, 2000a). In addition, given the multiple steps involved, and the requirement to maintain a goal state in mind, while working through sub-goal steps, working memory is a likely contributor, especially when the tasks are more difficult (Goel, Pullara & Grafman, 2001). As has been noted with other executive tasks, these tower tasks also seem to be sensitive to information processing speed in MS (Arnett, et al., 1997), though this sensitivity may relate to time being a key indicator of performance.

It is therefore likely that these Tower tests do reflect executive processes, but may be of limited use because of their gross and fine motor demands, and notably where time measures are used (e.g. would slow movement of a disc from one position to another be counted as thinking time or psychomotor speed). The issue of task purity is also of concern, given that simpler versions may require less in terms of working-memory and executive function than more complex ones. Computerised versions of the test are available, including norms, but selection and movement of the discs on the screen again may mean that this task is not appropriate for all because of visual and motor disabilities. A final consideration related to resources needed for this test to be used, especially access to both software and portable hardware. Tower tests were for these reasons not deemed appropriate to this study.

Probably the most frequently used test sensitive to executive function is the Wisconsin Card Sorting test (WCST; Grant & Berg, 1948). This task requires participants to sort sets of cards (each with differently sized and coloured shapes), where the sorting rule is not given and there are three possible sorting categories – size of geometric form, shape and colour. The only information given is either “right” or “wrong” feedback from the examiner after each card is sorted. The research participant therefore needs to use this feedback to generate the sorting rule. The sorting rule also changes without notice during the task. The test has been used in MS samples (Beatty, et al., 1995; Rao, Hammeke & Speech, 1987; Arnett, et al., 1994).

In a paper, which included a review of studies on the dimensionality and validity of the WCST as a measure of executive function, Royall et al., (2002) investigated what the dimensions of executive function might be, and the idea of the WCST as a gold standard measure of executive function. Generally the authors’ findings suggest the WCST might be variously associated with planning, concept formation and inhibitory functions, though this has been difficult to establish empirically.

In the Miyake et al., (2000b) study, *set-shifting* was proposed as the key demand of the WCST; this is the ability to move from one level of understanding (termed a ‘set’ - e.g. organise by shape) to another (organise by colour). It implies a level of cognitive or executive control to do the shifting, and may relate to the management of behavioural control, such as freedom from distraction, maintenance of task goal and of flexibility in task performance (Royall et al., 2002). Limitations of the task in consideration of this study relate to the length of time it takes to complete (about 45 minutes) and the potential impact of visual problems, including colour appreciation, upon completion. Finally, as discussed in the introduction, this is a ‘one-shot’ (Lezak, Howieson & Loring, 2004:588) test with previous experience likely to compromise reliability, making the second administration potentially more a test of recall, than executive function (Beatty, 1999). Another task, which has a similar structure of rule detection, rule change and a yes/no feedback to guide rule detection and change, is the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997).

The task is considered one of rule abstraction (Andres & van der Linden, 2000), concept formation (Lezak, Howieson & Loring 2004) or concept attainment (Burgess & Shallice, 1997). In essence, it requires the participant to divine the rule(s) governing the movement of a coloured circle around a 2 x 5 matrix of plain circles. The participant is told only that there are rules or patterns governing the movement and that these rules may change during the course of the task. Feedback is immediate in that the page is turned after prediction, and the participant can see whether they were correct or not in their prediction. A group of patients with circumscribed frontal lobe damage (‘Anterior and Bifrontals’ Burgess & Shallice, 1997:7) differed significantly from both normal controls and a group of people with posterior lesions on the mean number of errors made in the task, suggesting the task is sensitive to frontal lobe dysfunction. The test has been examined for construct validity alongside two more ‘standard’ tests of executive function (De Frais, Dixon & Strauss, 2006), using a confirmatory factor analytic approach. The authors suggest that all the tests loaded on a single executive factor.

For a control group, split-half reliability was 0.62, suggested to be adequate given the caveats about reliability in tests of executive function, and given what the authors term to be ‘serial dependence between trials’ (Burgess & Shallice, 1997; Burgess, 1997) that is, each choice depends in some way on the previous choices. The factoral relationship with other tests of executive function has been discussed.

As well as confirming that the Brixton test appears to be an appropriate index of executive function for this study, it supports selection of a second test – The Hayling Sentence Completion Test, also used in the De Frais, Dixon & Strauss (2006) study, as a potentially acceptable measure. Given suggestions of modality specific deficits in executive function

(Hunt & Kingstone, 2004), this second verbal test compliments the use of a non-verbal executive task (The Brixton test).

The Hayling Sentence Completion requires, in part 1, rapid completion of 15 sentences, with a single appropriate word, so as to finish the sentence in a sensible way. Part 2 again requires a single word to be given rapidly, this time to provide a nonsensical completion for a new set of 15 sentences. Effectively, part 1 of the test is designed to offer practice in automatic completion of everyday sentences. Then, part 2 requires inhibition of the 'trained' automatic completions so that the participant needs to inhibit the likely prepotent single word completion and generate an alternative, which is in no way related to the sentence; this is carried out under time pressure so as to be sensitive to where inhibition costs (a speed/accuracy trade-off) occur – longer response times or more errors in word generation. The authors, Burgess & Shallice (1997) describe the test as indicative of a number of executive-related abilities; response initiation and response suppression (measured in both number of errors or time taken to generate response; Burgess & Shallice, 1997:5).

De Frais, Dixon & Strauss (2006) demonstrated that both the Brixton and Hayling Sentence Completion tests loaded on the same factor with other measures of executive function, supporting their construct validity as measures of frontal or executive function. Additionally they demonstrated measurement invariance across gender and age, at least in older people (Young-old and Old-old adults; n=427 community dwelling adults, mean age 68.44 years). The Hayling Task was therefore selected as a second measure of executive function for this study.

Inter-rater reliability for this test may be as high as 96% (Bielak, Mansueti, Strauss & Dixon, 2006), despite the test allowing latitude on the interpretation of errors. The test manual (Burgess & Shallice, 1997) reports split-half reliability for the three component scores, for a control group, as r = 0.35 (sensible completion – time), r = 0.83 (nonsense completion – time) and r = 0.41 (nonsense completion - error score). For an impaired group with anterior (i.e. frontal) brain involvement the reliabilities are respectively r = 0.93, 0.80 and 0.72.

There is possibility for confounding results with information processing speed, given this is a timed task, and this will be considered in the light of planned testing of information processing as a mediator to cognitive abilities in the final set of model testing. As there is an ‘initiation’ (response time) score in the easy sentence completion task and again a response time score for the difficulty completion task, one possibility to reduced the confound is to subtract the easy condition response time from the difficult condition response time, to indicate the additional processing time. This could be considered the true executive time cost, rather than being confounded with response initiation because of information processing deficits.

A final approach to the assessment of executive function is the use of self-rating methods. Some self-rating instruments have been developed on non-neurologically impaired adults (Spinella, 2005), but are likely to be both unreliable and low in validity because they have not been assessed for their relationship with standardised measures, or for use in samples who have executive dysfunction (Bogod, Mateer & McDonald, 2003). These types of indices are not therefore considered optimal for this study; investigating the relationship between self-report of executive abilities and self-report of memory abilities, is likely to confound interpretation of results.

In summary, because there appears to be diversity in the subcomponents of executive functions, multiple assessments are typically indicated. In addition, given the traditionally low reliability of tests, multiple measures are probably warranted, and their factor structure will also be confirmed for this particular study sample.

Test selection was based on tasks that have limited complex visual and motor demands, and are not entirely speed based, or when they are, have a baseline speed measure for comparison. Finally, for the purposes of the study they are relatively quick to administer and have some demonstrated sensitivity for frontal lobe functions (Royall et al., 2002). The two tests selected to fit the range of criteria described - the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997) and the Hayling Sentence Completion Test (Burgess & Shallice, 1997), have an administration time of 15 minutes in total. Figure 2.3 summarise proposed latent variables selected for cognitive items.

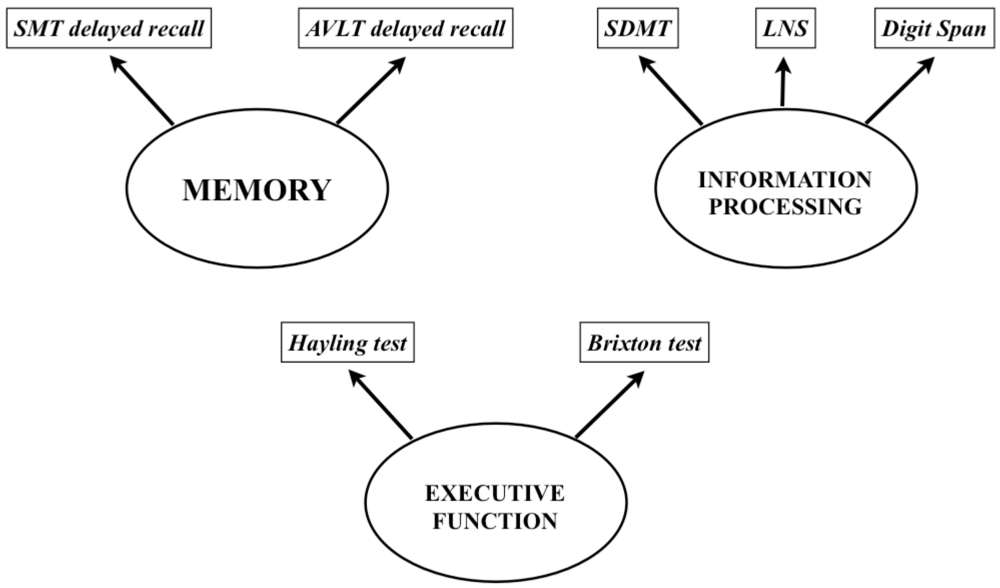


Figure 2.3 Summary of observed measures for the proposed latent variables; *Memory, Information Processing and Executive Function*

2.2.4 Selection of Measures of Mood.

Because of the purported association between depression and self-reported memory ability (Randolph, Arnett & Higginson, 2001; Randolph, Arnett & Freske, 2004, Bruce & Arnett, 2004; Zelinski & Gilewski, 2004), an assessment of mood is required for this study.

One key debate in the selection of measures of depression, or self-reported depression symptoms (Beck, Steer & Garbin 1988; Beck, Steer & Brown, 1996), relates to their potential sensitivity to neurovegetative symptoms (e.g. fatigue), considered in the Beck Depression Inventory (BDI) as indicators of depression when, in fact, they may be indicators of the neurological diagnosis (Nyenhuis et al, 1995). In other studies in MS (e.g. Randolph et al 2000), caution is advised in removing confounded items because these types of symptoms may be associated with both mood and neurological symptoms, or because some somatic symptoms may differentially load for depression in different individuals (Clark et al 1992; Aikens, et al., 1999; Bruce, McGuigan & Hutchinson, 2006; Polen & Arnett, 2007)

One issue discussed in Randolph et al., (2000), which used a revised version of the original BDI (Beck, et al., 1961; excluding tiredness, appetite and weight change) was the relationship between their version of BDI and the Chicago Multiscale Depression

Inventory (CMDI; Nyenhuis et al, 1995) mood scale. The CMDI Mood scale has been suggested to be a good indicator of depression in MS (Nyenhuis, 1995; Mohr et al, 1997), though the findings of Randolph, Arnett & Freske (2004), suggest that the 'constellation of depressive symptoms' assessed by the BDI may better indicate depression. The argument in some sense is moot if the newer conceptions of depression, with a more cognitive basis, are considered (American Psychiatric Association, 2000). Capturing this cognitive conception was the aim of revising the BDI more recently (Beck, Steer & Brown, 1996).

The issue of assessment of depression in MS has also been subject to expert review. The results published in 2005 (Goldman Consensus Group, 2005) supported the use of the BDI (Beck et al., 1961) for clinical assessment of depression in MS. It was accepted that the problem of MS-symptom and Depression-symptom overlap is a feature of screening assessments in general, including the BDI (Mohr et al., 1997; Goldman Consensus Group, 2005).

In a change from previous studies, the second version of the BDI (BDI-II; Beck, Steer & Brown, 1996) will be used. Two potential benefits are envisaged; the dropping of some of the items which may have been conflated with neurovegetative symptoms of MS – Somatic Preoccupation, Weight Loss, Body Image Change, and Work Difficulty. Second, it aims to reflect the updated American Psychiatric Association's definitions of depression (APA, 2000), reflecting a cognitive understanding of depression (e.g. the role of negative thinking styles), and considers mood over a two week period, not one. Factoral studies have proposed one, two and three-factor models based on the tool (Beck, Steer & Brown, 1996; Harris & D'Eon, 2008). In samples of people with 'fairly minor medical conditions' (Viljoen et al., 2003:289), more serious medical conditions (Thombs et al., 2008) and in primary care medical attendees (Arnau et al., 2001), a two-factor model Somatic/Affective & Cognitive has been proposed, or confirmed to be as good other two or three factor models (Thombs et al., 2008). All three studies supported the presence of a single second-order depression factor also.

In summary, the BDI-II (Beck, Steer & Brown, 1996) in its entirety will be used here. The tool takes about 10 minutes to complete and can be completed verbally, if necessary (Beck, Steer & Brown, 1996). The confirmatory factor analyses reported in Chapter 6 should clarify factor structure for this sample (1, 2 or 3 factors including affective, somatic or cognitive dimensions for example).

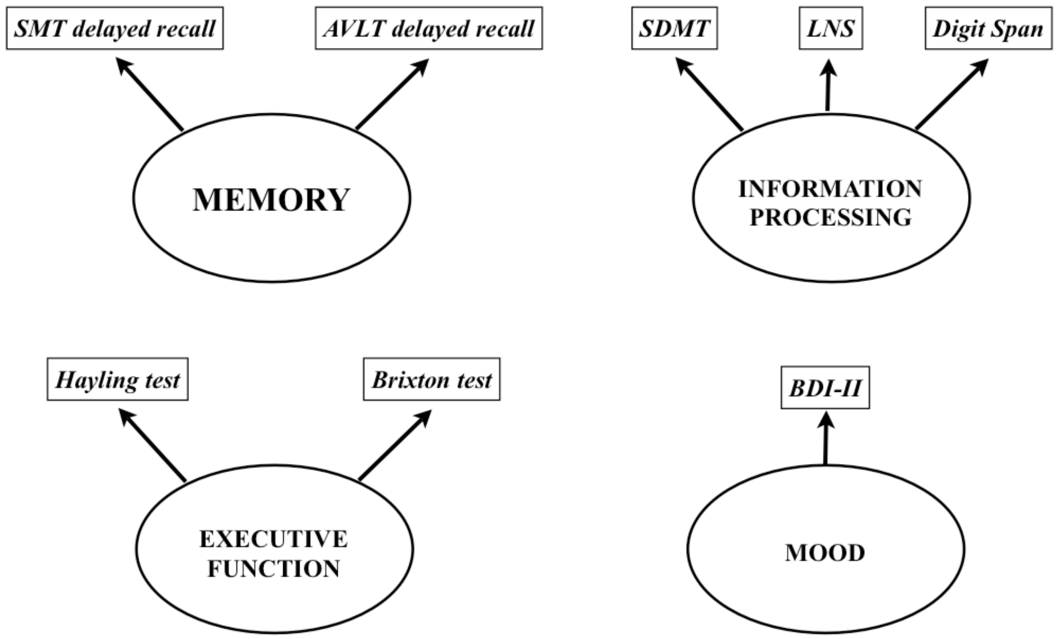


Figure 2.4 Summary of observed measures for the proposed latent variables; *Memory, Information Processing, Executive Function and Mood.* The 21 depression indicators of the Beck Depression Inventory will be subject to Confirmatory Factor Analysis to establish the number of factors to indicate the latent variable.

2.3. Selection of Metamemory Measures.

2.3.1 Introduction

Two main approaches to the assessment of metamemory monitoring have been discussed. The first was global assessment of memory ability, memory-self efficacy, or subjective evaluation of memory ability. The second approach is based on a number of on-line memory-related monitoring tasks.

A key assumption in the measurement of metamemory is that individuals act as their own measurement devices, in that introspection is the method by which access to internal ‘subject-controlled’ processes relating to memory is achieved (Nelson & Narens, 1980; Koriat & Goldsmith, 1996:491). There are a number of paradigms for assessment of these processes, each requiring some method of deriving or calculating accuracy, in order to make them meaningful (Nelson, 1984; Schraw, 1995; Moulin, 2002; Goldsmith, Koriat & Pansky, 2005; Pannu & Kaszniak, 2005; Spellman, Bloomfield & Bjork, 2008; Dunlosky & Bjork 2008). As a result of the relative subjectivity of the process of introspection, the measurement of metamemory does not have the same consensus about the calculation of accuracy as there is in quantity oriented memory assessment.

Generally, metamemory monitoring has been described as being either 'on-line' (made during task performance) or 'off-line' (made in respect of generally experienced memory; Veenman, Van Hout-Wolters & Afflerbach, 2006). One reason for the range of methods and analyses relating to metamemory performance may be that the tradition underpinning metamemory assessment differs from the approach to memory assessment in not being quantity-oriented (Koriat, Goldsmith & Pansky, 2000). Instead, it focuses on a correspondence view of memory, that is, the level to which, for example predicted recall corresponds to actual recall (Koriat, Goldsmith & Pansky, 2000). A second difference is that it does not focus on recall as its only index of memory-related processing; comparisons between judgements about ease of learning a piece of text, and time allocated to the task, is an example of other approaches to understanding metamemory in terms of the relationship between monitoring and control (Van Overschelde, 2008; Dunlosky & Bjork, 2008).

Monitoring processes are the primary focus of this study, because of the clinical experience driving the study. Also, because of the proposition that for people with neuropsychological impairment, defective monitoring is instrumental in inappropriate, or ineffective, control mechanisms (Moulin, 2002; Pannu & Kaszniak, 2005). Given this focus, and in consideration of the on-line and off-line measurement approaches, it is proposed that a spectrum of measures will be used. At one end, an 'off-line', task-independent measure that focuses on global subjective assessment of memory ability (memory-efficacy) will be sought. At the other end of the spectrum, are within-task predictive and postdictive measures of memory monitoring accuracy.

The on-line monitoring judgments focused on in this study are a delayed Judgment of Learning, a Retrospective Confidence Judgment and a Feeling of Knowing judgment. One aim in selecting the specific monitoring judgments is that they reflect monitoring at different stages of the memory process - Judgment of Learning during learning, Retrospective Confidence at retrieval and Feeling of Knowing prior to recognition, where retrieval has failed. The proposed mechanisms for each have been discussed in Chapter 1, and here the structure of testing will be discussed, followed by a review some approaches to the calculation of accuracy by correlation, proportion and probabilistic methods in Chapter 3: *Development of Statistical Methods.*

2.3.2. Selection of task-based Metamemory Measures.

Two types of memory tasks are used in metamemory research – one based on the use of sets of general knowledge questions (semantic recall; Hart, 1965; Shimamura & Squire, 1986). With semantic memory tasks, graded difficulty general information questions are available (Nelson & Narens, 1980; Shimamura, Landwehr & Nelson, 1981). Performance can therefore be compared to norms for both predictive ratings and accuracy of recall. Limitations of such lists relate to their applicability with the passage of time, with changes in education, or with changes in the facts themselves (i.e. the capital of Czechoslovakia). Grading of item difficulty is normed on university students, reducing external validity, as does the cultural biases inherent in such questions (e.g. the name of the 21st US president). The second approach is based on newly learned information, one benefit being that it may be more sensitive to monitoring deficits, especially for Feeling of Knowing (Souchay, et al., 2007).

Of the list of potential monitoring led measures discussed, those that minimise the possibility for variable lengths of time to complete (so as to avoid fatigue related confounds affecting some, but not other, participants) were sought. Measures where participants are allowed to control their own learning rate by definition will introduce variability in timescales due to different learning rates (Son & Kornell, 2008). This, along with the range of other tasks in the study, is likely to increase the time per participant also.

The proposed advantage of Judgments of Learning, Feeling of Knowing Judgements and Retrospective Confidence Judgements is that they can be built around a memory task without biasing the learning by drawing attention to the later recall component (unlike the judgements of ease learning perhaps). Therefore the measure can also be used as a relatively uncontaminated measure of recall. Only small additional time is required in the testing session, and they address different components of the memory experience -

encoding, recall and recall failure. Additionally there are measures that are traditionally reported in the metamemory literature in clinical samples.

Benjamin & Diaz (2008) present a typical paradigm for metamemory measurement, which includes three components; a manipulation of study or judgement conditions, a measure of metamemory and a test of memory. The relevant memory tests have been presented earlier; the Auditory Verbal Learning Test and a Sentence Memory Test and the considerations for selecting them with this population have been discussed. Here, the study or judgment manipulation for each is presented.

2.3.2.1. Judgment of Learning

The Judgment of Learning (JoL) is to be derived from the Auditory Verbal Learning Test, for which a global judgement about likely recall can be made on the delayed recall of this 15-item word list. As discussed, the task involves the serial learning of a 15-word list to which the participant is exposed 5 times. Then an interference list of 15 different words and a delay of 20 minutes provided, before recall of the original list is tested. Prior to offering each trial of the list, including the first, the participant will be asked to estimate how may of the 15 items they are likely to recall.

While not fully in line with the typical structure of Judgments of Learning, where participants estimate whether they have learned something on an item-by-item basis for word pairs (Arbuckle & Cuddy, 1969), the proposed structure allow for factors that influence JoL accuracy - number of study trials and the timing of making the JoL - immediate or delayed (Kelemen & Weaver, 1997; Dunlosky & Metcalfe, 2009). This is a more global judgment about learning than has traditionally been used, but perhaps reflects more validity when considered in a real-life context. The aims of including a Judgment of Learning such as this are threefold. It focuses on a monitoring judgment relating to learning. Secondly, it allows for investigation of sensitivity to learning, a factor that may be relevant in people with memory impairment (Moulin, 2002). Finally presence of the underconfidence-with-practice can be determined.

2.3.2.2. Retrospective Confidence Judgment.

In the Sentence Memory Test, once recall for each of the 24 sentences is tested, participants will be asked to provide an indication of their confidence in the answer that they have recalled (Retrospective Confidence Judgement, or RCJ; Pannu & Kaszniak, 2005). The key measures relate to the association between confidence rating and correct recall, which can be considered both in terms of proportional accuracy and relative accuracy in the association of confidence with ‘correctness’. These two measures reflect calibration and resolution in judgment accuracy (Dunlosky & Metcalfe, 2009). For the purposes of investigating both accuracy and inaccuracy, it is proposed that for proportional measures, both proportion of High confidence/recalled and Low confidence/recalled will be examined. While the High Confidence/recalled indicates accuracy, a judgment of Low confidence for successfully recalled items is of interest in the context of cognitive and affective disturbance. It might imply better than sensed availability of a memory trace, or indicate a tendency towards negative appraisal.

In summary, relative accuracy in RCJ will be measures and two indicators of absolute accuracy - accuracy and inaccuracy in calibration. The measurement approaches will be discussed in mode detail in the following chapter.

2.3.2.3. Feeling of Knowing

Developed from the work of Hart (1965) and used by others with neurologically impaired populations (Shimamura & Squire, 1986; Beatty & Monson, 1991; Souchay, Isingrini & Gil, 2002; Souchay, Isingrini & Gil, 2006), the Feeling of Knowing task asks participants to rate the likelihood of recognising currently unrecallable information, using a scale of ratings or rankings. The unrecallable information is derived from performance on a recall task (here, the Sentence Memory Test), and selecting items for which no answer or an incorrect answer was offered. Prior to a recognition trial, participants are asked to rate the chances of recognising the correct answer when they see it. Accuracy of the judgment is typically assessed in both absolute and relative terms, discussed further in the following chapter. Figure 2.5. summarises the task structure and indicates one approach to determining accuracy, by correlating judgment and performance.

It is recognised that there are some challenges in asking people to grade their Feeling of Knowing (Beatty & Monson, 1991). The aim will be to have each participant rank each of their unrecallable sentences so as to indicate the relative strength of their FoK, highest to lowest and then proceed to a recognition trial, which will be a 7-alternative forced choice test among plausible lures, based on responses given by a sample of people from the UK (Arcuri, et al., 2001).

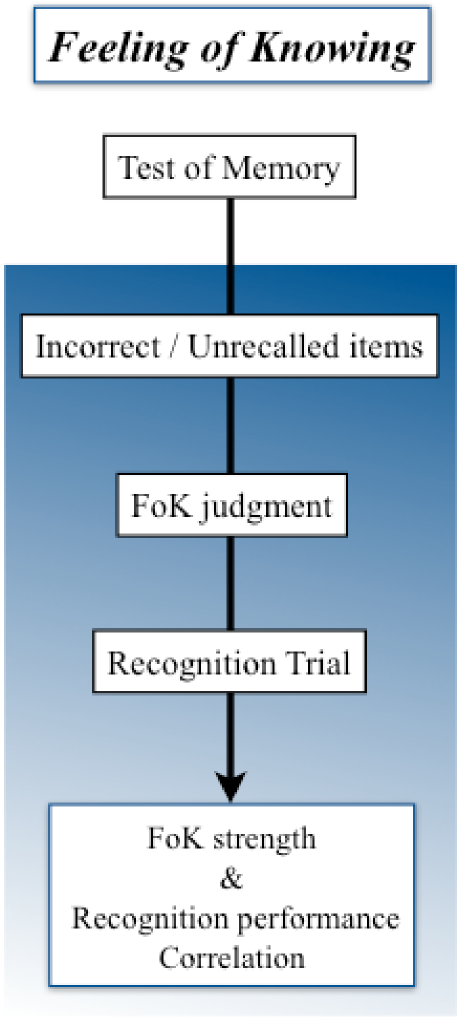


Figure 2.5. Structure of the Feeling of Knowing (FoK) task.

It has been established in the literature that these indices of metamemory reflect differing demands, so it would not be expected that they would be factorially related (Dunlosky & Metcalfe, 2009; Leonesio & Nelson, 1990) and will therefore, in contrast to the cognitive and mood items, not be pursued for their factorial convergence in confirmatory factor analysis.

2.3.3. Memory Self-report Measure.

This approach to the measurement of metamemory, in which global judgements about memory performance are made, is often termed self-reported memory, memory complaint, subjective memory appraisal or memory-efficacy (Hertzog, 2002; Zelinski &Gilewski, 2004). Given the weak or absent correlations between questionnaires about and performance in memory tasks, the reliability and internal consistency of questionnaires has been questioned (Hertzog, 2002).

Two questionnaires in particular have been closely examined in the literature – the Memory in Adulthood (MIA; Dixon, Hultsch & Hertzog, 1988) questionnaire and the Memory Functioning Questionnaire (MFQ; Gilewski, Zelinski & Schaie, 1990). Randolph, Arnett & Freske, (2004), used the latter MFQ scale with their MS sample. Since one objective in this study is to look at the stability of their proposed model in a different MS sample, it was selected for use.

The main alternative to the MFQ is the Memory in Adulthood questionnaire (Dixon & Hultsch, 1983; Dixon, Hultsch & Hertzog, 1988), consisting of either 108 or 120 questions (Dixon & Hultsch, 1984), scored on a 5 point Likert scale. The length of the MIA would potentially detract from its use here, as would the limited 5-point response scale. Additionally, Cavanaugh & Green (1990), summarising the study of Hertzog, Hultsch & Dixon (1989), supported the Frequency of Forgetting scale, part of the MFQ, as having a high level of convergent validity with a similar efficacy factor in the MIA. In this regard, the selection between them becomes less an issue of contrasting validity and more an issue efficient data collection.

Despite misgivings about the technical value of questionnaires about memory performance, they form an essential part of this study for a number of reasons. Their ongoing use in clinical assessment, previous studies using them to investigate metamemory in MS (Beatty & Monson, 1991; Randolph, Arnett & Higginson, 2001; Randolph, Arnett & Freske 2004) and finally for comparison with on-line measures of metamemory. A reasonable summary, in respect of questionnaires or memory complaint, would suggest that affectivity, implicit beliefs, conscientiousness and age relate to how questions are answered (Hertzog, 2002; Zelinski & Gilewski, 2004). Some studies have proposed that recall independently explains a small amount of variance in efficacy scales (Zelinski & Gilewski, 2004).

2.3.3.1. The Memory Function Questionnaire

The questionnaire selected for use in this study is the Memory Function Questionnaire (Gilewski, Zelinski & Schaie, 1990; Zelinski, Gilewski & Anthony-Bergstone 1990), proposed by the authors to measure self-perception of memory functioning. Minimum scoring is 64, maximum is 448, assuming all questions are answered. It is further proposed that the factor structure is age-invariant across two age groups (16-54years and 55-89 years), and that it has high levels of internal consistency for each factor (all > 0.80). In one study, age did account for some variance in overall scores, with older subjects endorsing more memory difficulties (Gilewski, Zelinski & Schaie, 1990), as did levels of education and general health. The authors caution over-interpreting this as the total variance accounted for by all three was less than 9%. In another study, some evidence is presented of a correlation between scoring on this questionnaire and tested memory performance (Zelinski, Gilewski & Anthony-Bergstone 1990). At 64 items, the scale is likely to have some redundancy included, so further investigation, with a view to reduction was carried out.

Principal component analysis has suggested four factors, the first of which contained 33 of the 64 items in the questionnaire and was interpreted as a General Frequency of Forgetting factor (Gilewski, Zelinski & Schaie, 1990). This index incorporates questions about general frequency of memory lapses in everyday situations, frequency of forgetting while reading, questions about ability to recall over selected periods of time and an overall rating of memory ability. Subsequent analysis of this factor (Zelinski & Gilewski, 2004) proposed that it could be further reduced to a 10-item scale, through a Rasch scaling process. Variance of the 10-item General Frequency of Forgetting scale was explained by conscientiousness, depression and list recall, independent of the other variables (Zelinski & Gilewski, 2004). As one of the aims of this study is to compare previous use of this instrument, the entire 64-item measure, with a 7-point Likert response scale will be used. Model testing will focus on the Forgetting While Reading subscale to comparing with the findings of Randolph, Arnett & Freske (2004). This scale is based on the selection of questions of reading related questions, and is not formally a subscale of the Memory Function Questionnaire. In addition, the study will use the 10-item General Frequency of

Forgetting scale. This scale, based on its statistical derivation, is proposed to be a more robust measure with which to index memory efficacy. Figure 2.6 summarises the metamemory measures to be used in the study.

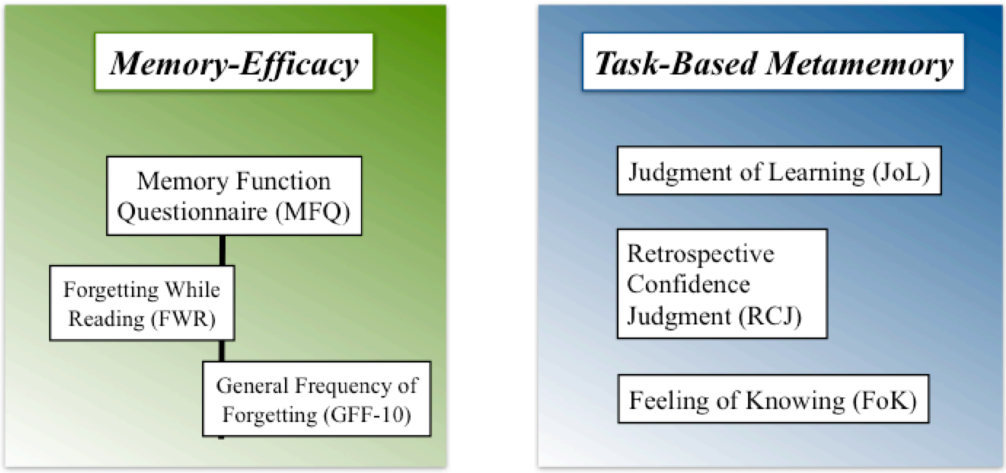


Figure 2.6 Summary of Metamemory measures to be used in the study.

2.4. Summary

A number of tests have been reviewed in order to select the most appropriate indicators of the key variables for the study. These tests have been selected based on knowledge of their reliability, validity and appropriateness for the sample and processes in this study.

The development of latent variables, constructed from multiple indicators selected in this chapter requires testing for factoral coherence however. In the next chapter the statistical considerations for this process - Confirmatory Factor Analysis (CFA) is considered. The methods discussed with also apply to the testing of relationships between latent variables themselves, the structural equation models. Additionally, approaches to quantifying accuracy in metamemory performance and underpinning statistical concerns will be discussed. These include sample size, missing data management and assessment of data quality.

Chapter 3: Development of Statistical Methods.

3.1 Introduction.

Three sections of methods of analysis are considered in this chapter. The first will relate to analysing performance on the range of tests used, required to summarise characteristics of the samples' performance. This analysis will be used to describe the sample characteristics, in order to situate performance on cognitive, affect and metamemory measures in relation to normative levels. To this end, clarification of what will be considered impaired performance is first considered in this chapter. This will be followed by a consideration of how accuracy in metamemory measures is approached, the sample size and approaches to the assessment of data quality.

The second set of methods of analysis, the *Measurement Model*, will explain assessment of the observed variables in terms of their factoral structure. The aim is to describe congeneric observed variables that can be used to create latent variables for each of the domains under investigation when the structural models are tested. The method of testing measurement models is Confirmatory Factor Analysis and statistical methods for this analysis are outlined. The final set of methods of analysis presented is the testing of latent variable or *Structural Equation Models*, needed to address the central objectives of the overall study. Statistical methods for the testing structural models are broadly similar to those of the measurement models but will be considered in respect of the additional elements relating to mediating processes in structural models, and how that is assessed.

3.2 Defining Impairment in Selected Measures

The decision about placing a point at which a score is considered abnormal is a trade-off between detecting impairment in those who are impaired and excluding those who are not, so-called specificity and sensitivity (Daly & Bourke, 2000; Sbordone, Saul & Purisch, 2007) The setting of a cut-off score is also made in recognition of the many different conceptions of what normal is - clinical (where average or above scores are not of concern), prognostic or statistical (Daly & Bourke, 2000).

For the purposes of evaluation of the sample, the distinction between impaired and unimpaired, will be considered as at or below the -1.5 standard deviation (SD) mark (7th percentile). This cut-off lies between 1 standard deviation (16th percentile) and 2 standard deviations (approximately the 2nd percentile) below the mean score. This is a common point in many neuropsychological tests (Smith, 1982; Sbordone, Saul & Purisch, 2007), that avoids an overly 'sensitive' -1 SD (16th percentile) and a potentially too 'specific' -2 SD (2nd percentile) in a normal distribution of performance (Field, 2005). This point will be used for all tests in conjunction with published norms.

For the Beck Depression Inventory 2nd Edition, categorical descriptors have been set by the test authors, which reflect a relative emphasis on sensitivity over specificity (Beck, Steer & Brown, 1996). These categories also aim to indicate the severity of reported symptoms.

For metamemory, the focus will be on accuracy of judgment, rather than impairment. There have been few studies on metamemory in MS, and those based on neurologically-normal samples have used a variety of measurement tasks, making comparisons difficult. The statistical approaches to accuracy measurement are therefore now considered.

3.3. The Measurement of Accuracy in Metamemory.

Two opinions bear on the selection of metamemory measurement approaches in this study. Nelson & Narens' (1980:70) recommendation that *'it is desirable to employ sound techniques that have a relatively straightforward interpretation'*, and Irwin's suggestion (1934 *cited in* Goodman & Kruskal, 1959:156) that:

> *'we should try to make the end point of the statistical analysis not a single coefficient which may be hard to interpret, but a result bearing a 'physical' meaning; the more easily the result may be understood by an intelligent layman, the better we should regard it expressed'.*

There remains considerable debate about the measurement of accuracy both philosophically (e.g. Kruglanski, 1989), methodologically (Benjamin & Diaz, 2008: in Dunlosky & Bjork) and statistically (Nelson & Narens, 1980; Nelson et al., 1986; Schraw, 1995; Nietfeld, Enders & Schraw, 2006; Veenman, Van Hout-Wolters & Afflerbach, 2006; Spellman, Bloomfield & Bjork, 2008).

A number of measures of metamemory have been selected for this study, with the Memory Function Questionnaire and its Forgetting While Reading, General Frequency of Forgetting and total scores all derived through summation of the grades given by each participant. The accuracy of this 'off-line' measure will be investigated in relation to performance on memory tests. Only on-line measures of metamemory, from which measures of accuracy can be derived, are therefore considered here.

These are the *Judgment of Learning* (JoL), *Retrospective Confidence Judgment* (RCJ), the *Feeling of Knowing* (FoK) judgment. A description of each in terms of its measurement structure has been presented previously. All three are in some way measured in respect of how the predicted performance or confidence in performance relate to actual performance. Narens (2002) presents a discussion addressing a justification for the assignment of numerical values to subjective experience and thence using common 'classical' measurement approaches to analyse these numerical values. The conclusions are that these assigned values (given to subjective experience such as 'more confident') can be treated as *'numbers'* (Narens, 2002:787). Consideration is warranted of measures of accuracy that do provide interpretable data, as well as reflecting the range of accuracy measures traditionally use to allow for some comparison of results. A final factor in deciding the accuracy measures is pragmatic; what the different tests can yield in terms of measurable accuracy, and what might be of clinical utility in neurological populations. The selection of these measures is now outlined.

3.3.1. Relative Accuracy, or Resolution

Relative accuracy, or *resolution* (Dunlosky & Metcalf, 2009) is typically correlation-based, such that a correlation measure is derived based on each individual's judgments and their performance, that is, a within-subject measure (Benjamin & Diaz, 2008). As a correlation, it indicates that a participant *'can discriminate between the differences in the memorability of the items'* (Dunlosky & Metcalf, 2009:49) or predicts their performance on one item relative to another (Dunlosky, Rawson & McDonald, 2002). Koriat & Goldsmith (1996) suggest that this item-by-item assessment of the probability of being correct is the monitoring process used in free recall tasks. This has implications for the types of measures that are selected; that they reflect a correlation and that they are structured so that a participant can indicate a higher or lower strength in the judgment e.g. in Feeling of Knowing or Retrospective Confidence (Nelson, 1984; Koriat, Goldsmith & Pansky, 2000; Benjamin & Diaz, 2008; Dunlosky & Metcalf, 2009) It also has implication for how resolution is calculated - it needs to cope with a range of strength predictions (2+; Nelson, 1984; Nelson & Narens, 1990) and performance outcomes (typically Yes/No indicators for recall or recognition).

As the Feeling of Knowing and Retrospective Confidence Judgments planned provide the opportunity for item-by-item judgment making, this would mean that the strength of feeling of knowing an answer should relate to the likelihood of correctly selecting the answer; for retrospective confidence, the level of confidence should relate to the likelihood that the given answer is the correct one. Many researchers (Davis, 1967; Nelson, 1984; Beatty & Monson, 1991; Nelson 1996; Dunlosky, Rawson & McDonald, 2002; Benjamin & Diaz, 2008), though not all (Schraw, 1995; Pannu & Kaszniak, 2005), have proposed that a non-parametric gamma correlation (γ, or Goodman-Kruskal correlation; Goodman & Kruskal, 1954; Goodman & Kruskal, 1959; Nelson, 1984; Nelson & Narens, 1990) is the most appropriate, or least worst (Spellman, Bloomfield & Bjork, 2008), measure of the relationship between predicted and actual performance.

The relationships between Retrospective Confidence Judgments or Feeling of Knowing judgments and actual performance are each viewed as a bivariate association indicating *'a within-subject correlation of performance and judgement'* (Benjamin & Diaz, 2008:76). As with standard parametric correlations, values range from -1 to +1 (Goodman & Kruskal, 1954; Goodman & Kruskal, 1959) giving it high levels of interpretability. Being easily understood, it therefore has some potential utility in terms of the regulation of learning (Dunlosky, Rawson & McDonald, 2002) because of *'good psychological transparency'* (Benjamin & Diaz, 2008:77; Nelson, 1984). The indication of relative strength is commonly carried out by a rating or ranking of items, according to the strength of the feeling, and comparing these to performance.

One consideration prior to investigating methods for the calculation of relative accuracy is the difference between rating and ranking methods of indicating judgment strength. Rating-based approaches, which might use a set of 4 ratings create, categories of judgment strength, allowing for only ordinal and nominal levels of analysis (Rating category and Yes, No response). Rank-based methods generate a single ranked judgment list from *highest* to *lowest*, allowing for an ordinal (ranks) to nominal (Yes, No) analysis. Both approaches mean the calculation of a non-parametric Goodman-Kruskal gamma correlation (*G* or γ) is warranted (Goodman & Kruskal, 1954; Nelson, 1984).

3.3.2. Goodman Kruskal Gamma. (Goodman-Kruskal, 1954; 1959; Nelson 1984, 1996; Nelson & Narens, 1990; Schraw, 1995).

This is a classic approach to calculation of *relative* accuracy in that it provides 'inferences about bivariate association' (Gibbons 1993). More simply put, it describes relative accuracy, of ordered variables (Schraw 1995), in terms of the association or the 'degree to which predicted performance on one set of variables corresponds to actual performance on the same set of variables' (Nietfeld, Enders & Schraw 2006 pg 260). It is considered equivalent in interpretative terms to other correlation coefficients, such as the Pearson (Beatty & Monson, 1991) and ranges in value from +1 to -1. For the ratings-based judgments items with a high judgment 'strength' (e.g. feeling of knowing) and a positive outcome are considered concordances because the level of subjective judgment is associated with the same level of performance; likewise low judgment strength, associated with low performance, is also considered a concordance.
Based on a 2 x 2 table of results, see table 3.1

Table 3.1: A 2x2 matrix of possible responses in a judgment task cross-tabulated with Recognition/Recall performance, where Yes' indicates correct recall/recognition. Cells *a* and *d* are concordant because strength and performance associated, whereas cells *b* and *c* are considered disconcordant because judgment strength and actual performance are not in agreement.

		Recognition/Recall Performance	
		Yes	No
Judgment Strength	High strength	**a**	**b**
	Low strength	**c**	**d**

Both *a* and *d* would be considered concordant responses, and *b* and *c* disconcordant in that the judgments align with performance in the former, not in the latter.

Using the rating procedure, the calculation of gamma is given alternatively by Schraw (1995) as

$$\frac{a.d - b.c}{a.d + b.c}$$

fig. 3.1

by Beatty & Monson, (1991) as;

$$\frac{\#\,\text{concordances} - \#\,\text{disconcordances}}{\#\,\text{concordances} + \#\,\text{disconcordances}}$$

fig. 3.2

and by Spellman, Bloomfield and Bjork (2008:105) as;

$$\frac{\text{concordances} - \text{disconcordances}}{\text{concordances} + \text{disconcordances}}$$

fig. 3.3

It is not clear from the published research whether the number of concordances is achieved by summation or is the product of concordant cells in the contingency tables, though some have presented the gamma derived by summation as the Harmann coefficient, not the Goodman Kruskal gamma (Schraw 1995; Nietfeld, Enders and Schraw, 2006). For this study therefore, the method provided by Schraw (1995) will be used. If necessary, this method also proposes collapsing the four ratings to a *High* or *Low* judgment, reflecting a consideration that the judgment be scaled the same as the performance criterion (Recalled/Recognised or Not recalled/Recognised; Dunlosky & Metcalfe, 2009; Koriat, 2002).

Some difficulties therefore emerge in using rating and ranking processes, notable where the complete set of rating options is not used or where complexities arise in the ranking of all items according to judgment strength (Nelson & Narens, 1990). Both are therefore considered.

3.3.2.1. Rating-based approaches

Some studies have used a 4-point rating scale, with a rating of 4 indicating a high level of confidence or feeling of knowing for an item, and 1 reflecting very low levels of confidence or feeling of knowing. For the purposes of clarity the following discussion will relate only to Feeling of Knowing, as the methods are similar for RCJ.

Using this rating approach, the calculation of a FoK gamma correlation from the Sentence Memory Task can allow for two possible contingency tables - one with the four ratings of Feeling of Knowing (4=high, 3=medium, 2=low and 1=none) and the two performance outcomes (recognised or not-recognised). This will yield a 4 x 2 table for gamma calculation. A second approach is to retain a 2 x 2 matrix in which the same two performance outcomes are used, but only two ratings of strength of Feeling of Knowing are calculated - High or Low; scores of 4 and 3 are considered ‘high’ and scores of 2 or 1 ‘low’.

Beatty & Monson (1991) proposed an approach collapsing ratings into a High and Low category if the refined ratings were too difficult or if the full range of the scale was not used. If only ranks 3 or 4 were used, for example, 4 was given a ‘high’ value and 3 a ‘low’ value. This phenomenon of restricted rank use has been reported in other neurologically impaired groups (Moulin, 2002), though it may be related to participants having more severe memory disorders, rather than task difficulty per se. In truth the two may not be separable, as episodic memory tasks may be difficult to those with episodic memory deficits (Schwartz & Metcalfe, 1994; Pannu & Kaszniak, 2005). For this study, the four-rating procedure will be used initially for both FoK and RCJ judgments and a gamma correlation will be calculated from this. Appreciating that rating non-use may be a feature of the more difficult Feeling of Knowing judgment task in clinical samples, use of the full ranking procedure will also be carried out.

3.3.2.2. Ranking-based approaches

The four-ratings procedure for assessing judgment strengths has been discussed. As well as an end in itself for organising relative strength of a judgment, it can also be used as a starting point for ranking judgment strength in an item-by-item way, especially for the FoK judgment (Shimamura and Squire, 1986; Nelson & Narens, 1990; Beatty & Monson, 1991). Unlike the rating-based procedure this will give an ordinal/nominal table from which the Goodman-Kruskal gamma correlation can again be calculated. Since the cross-tabulation calculation discussed earlier typically ignores ties, this procedure might better reflect of performance. As ties cannot occur because only one ordinal rank can be used for any one item, more of the judgment and performance data is therefore included in the calculation than in the rating-based approach. Table 3.2 gives an indication of the differences between the two methods, based on six notional FoK judgments. The first column indicates the rating procedure, column two the ranking procedure, and column three the performance.

Table 3.2: Comparisons of Ratings-based and Rankings-based assessment of Feeling of Knowing judgment accuracy. Scores of 1 to 4 in the *ratings* of judgment indicate categorical strength of judgment strength e.g. *High, Medium, Low* or *'pure guess'*; scores in the *rankings* indicate ordinal strength of judgment since each rank can be used only once thereby indicating relative strength of the judgment compared to each other items' judgment. Performance is binary in that an item is either recognised or not.

Rating	Ranking	Performance Recognised = 1 Not Recognised = 0
4	6	1
4	5	1
3	4	0
3	3	1
2	2	0
1	1	0

In summary, two derivations of the Goodman-Kruskal gamma correlation are available one based on *rating* of judgment strength, the other based on *raking* of judgment strength. The rating-based approach will be used for both Feeling of Knowing judgment and Retrospective Confidence Judgment in the first instance. Bearing in mind that for the Feeling of Knowing judgment, rating-non-use may be a feature of the judgment making, the ranks-based approach will also be sought to allow for a gamma correlation to be generated also. Retrospective rating of judgment strength for the confidence judgment is unlikely to pose the same difficulties based on the limited literature in neurological samples (Kennedy, 2001). It may be inappropriate to seek a ranked list of confidence in already retrieved items as participants, having retrieved 24 items may be unable to discriminate their relative confidence, as a high number at that stage will be known.

3.3.3. Absolute Accuracy, or Calibration

Unlike relative accuracy, absolute accuracy or calibration provides insight into the degree of agreement between predicted versus actual performance; that is, difference and proportion scores (Hacker, Bol & Keneer, 2008). As such, under and over-confidence are relevant considerations, the implications of which have been found to be important to performance in educational settings (Maki & McGuire, 2002; Hacker, Bol & Keener, 2008).

A number of indicators of absolute accuracy are available, variously based on differences between predicted and actual recall or proportional performance for judgments (e.g. Harts difference score, Hart, 1965; Nelson 1984). Measures of absolute accuracy reflect how accurate prediction is compared to actual performance (Koriat, 2002; Nietfeld, Enders & Schraw, 2006; Dunlosky & Metcalfe, 2009). This means that interpretation is based on, for example an 'out of 10 items how many were correct and how many incorrect', and how does this relate to the proportion predicted to be correct.

Often these measures derive accuracy by the subtraction of the incorrect from correct, as a proportion of all decisions. However, in clinical samples it is of interest to consider differences between inaccurate and accurate calibration to assess whether the processes related to being accurate or inaccurate are different, as they might be with different types of cognitive impairment. One such instance might be in those with poor recall but high confidence (Johnson et al., 2000; Moscovitch & Wincour, 2002). Such a separation has not typically been the focus in previous studies. Instead, performance is often amalgamated into a single calibration score reflecting proportion correct minus proportion incorrect, as the measure of accuracy.In keeping with the proportional focus of these judgments, and with discrepancies between prediction and actual performance, these measures will therefore focus on proportion of each judgment correct, both for high and low condition (High FoK, Low FoK; High RCJ, low RCJ) in an attempt to understand both accurate and inaccurate calibration.

For the delayed Judgment of Learning, a discrepancy score will be calculated as predicted minus actual performance on the 15-word list after a delay.

3.3.4. Summary

In line with much of metamemory research, measures of absolute and relative accuracy will be derived. For relative accuracy the measure used will be the Goodman-Kruskal gamma - based on both rating-based judgment and ranks-based judgment for the FoK task, on rating-based judgment alone for the RCJ. For absolute accuracy - proportional measures and a discrepancy measure will be used. For delayed JoL, a discrepancy between predicted and actual recall, for FoK and RCJ measures, focusing separately on *high judgment strength - high performance* (accuracy focus) and *low judgment strength - high performance* (inaccuracy focus).

The remaining sections of this chapter will review statistical considerations in respect of sample size and data quality, ending with a review of the statistical methods associated the two-step process in Structural Equation Modeling - the *Measurement Model* and the *Structural Model.*

3.4. Sample Size

The issues relevant to ideal sample size in Structural Equation Modeling (SEM) relate to both components of the process - confirmatory factor analysis and structural model assessment. Various heuristics exist suggesting 100-150 cases in CFA models with only two indicators per factor (Marsh & Hau, 1999; Kline, 2005), or a minimum for structural models of 100-150 cases (Ding, Velicer & Harlow, 1995; Schumacker & Lomax, 2004; Boomsma, 1982; Marsh & Hau, 1999). Other recommendations range from 5 to 10 cases per latent variable, or per estimated parameter, in a structural model (Bentler & Chou, 1987; Kelloway, 1998). One additional benefit of larger samples, is in providing cases for cross-validation of developed models (Schumacker & Lomax, 2004). Generally, given the global recommendation for about 100 as a minimum sample size, this study therefore aimed to recruit 100 participants.

Small sample sizes tend to lead to acceptance of models (Type II error), where large samples tend to lead to rejection of models (Blunch, 2008). As a result there are a range of ways to judge acceptability of models aside from the primary measure - the x^2 statistic and its associated p-value. The x^2 statistic and the range of other model fit indices are discussed later in this chapter; some are more prone to the influence of sample size than others. It is therefore advised that a range be reported to best characterise a model’s fit with the data (Bollen, 1990).

Schumacker & Lomax (1996) offer some suggestions to minimise the issues associated with sample size; use of sound theory, set out a-priori models, use the two step approach so that each latent variable is tested and confirmed before the relationships between them are confirmed. In addition to reflecting the psychometric properties of the measures used, many of the issues of sample size also relate to the distribution of data, amount and treatment of missing data and. A consideration of the data distribution issues, including missingness handling is therefore now presented.

3.5. Missing Data

The goal of missing data evaluation and treatment is to allow better inferences to be made about the population of interest, rather than having a complete dataset from which to make estimations or derive the results that would have been seen with a complete dataset (Schafer & Graham, 2002). Graham (2009:509) addresses the questions about the function of missing data analysis and treatment from the point of view of ‘preserving’ the characteristics of the dataset as a whole. Schafer & Graham (2002) argue for example, that while filling the gaps in variables’ data with the average score in the variable (mean substitution) might accurately predict individual missing values, it will as a result, reduce both variances and covariances (Byrne, 2001) and bias the relationships between the data when estimating the population-based inferences that can be made.

One of the key issues with the management of missing data is an understanding of any patterns to the missingness (Rubin, 1976; Allison, 1987; Little & Rubin, 1987; Byrne, 2001; Graham, 2009) in order to support a ‘statistically principled’ approach to its management (Wayman, 2003:3). Patterns of missing data can relate to problems with the instruments being used, the methods in which data is gained, or specific features within the population under study (Roth, 1994; Schafer & Graham, 2002). Questions about the extent to which there are relationships between gathered and missing data, or one piece of missing data and another, have led to the development of a taxonomy for describing patterns of missingness (Little & Rubin 1987, 1989; Graham, Cumsille & Elek-Fisk, 2003). The first type of missingness is data Missing Completely At Random (MCAR); this is data that is missing because of a random event such that a coin toss could be the mechanism (Graham, Cumsille & Elek-Fisk 2003). The cause of missingness is not associated with the variable containing missingness. Missing At Random (MAR: Little & Rubin, 1987) data is where the cause of the missingness is not due to the missingness variable itself, but it is missing conditional on another variable (Schumacker & Lomax, 2004), so that the reason for missingness could be linked to other variables in the data

(Byrne, 2001). The *random* here therefore means that the missingness, once the other variables in the dataset are controlled-for, does not depend on some unobserved variable; it is random (Graham, 2009).

A final missingness mechanism relates to situations where the reason for missingness is Non-Ignorable (Little & Rubin, 1987); it is Missing *Not* At Random (MNAR: Graham, Cumsille & Elek-Fisk 2003) in the sense that it is systematically missing in some way (Byrne 2001) related to the data collection mechanism, the tools or the variables under study. Understanding patterns to missingness is important because of their potential effects on the estimations about the populations under study. Data missingness that is a random component will not typically bias estimates about the population; with MCAR it is 'random'. With MAR it is 'random' only if the cause of the missingness is taken into account. MNAR, because it is not random, will yield biased estimates, relating to the reason for the missingness (Graham, 2009).

Wiggins & Sacker, (2002:106) discuss the approaches to managing missing data as a continuum from 'ad hoc' approaches, through simple model-based solutions to more complex model-based solutions. Approaches are a continuum from deletion approaches (listwise, pairwise), single imputation (mean and regression), Multiple Imputation and Maximum Likelihood (ML) estimation approaches. The latter two approaches come more generally recommended (Wiggins & Slacker, 2002; Schafer & Graham, 2002; Graham, 2009)

3.5.1. Deletion approaches

In general there is a range of treatments for missingness that are considered inappropriate, for various reasons. These include methods that effectively 'pretend there is no missing data' (Graham, Cumsille & Elek-Fisk 2003:90) and include deletions - such as listwise (whole case deletion) and pairwise (item deletion). One major limitation of the listwise deletion approach is that there may be differences between the people who complete tasks, and those who do not. Deleting the data from those who do not fully complete a questionnaire for example is likely to bias the estimation toward those who complete the questionnaire; people who fail to complete questions about memory performance may do so because of problems with memory. Deletion of the whole response set of individuals who do not complete a task is likely to bias the population estimated performance, perhaps by deleting incomplete data from more impaired participants. An additional limitation of this procedure is the potential for significant losses in sample size (Graham, 2009).

Unlike listwise deletion, pairwise deletion only deletes items where they turn up as missing data. This means that, on the face of it, as much data as possible is being used in the covariance matrices that are the foundation of the SEM process. However, it also leads to statistical issues relating to differing samples sizes for different elements of the covariance matrix. (Graham, Cumsille & Elek-Fisk 2003), and can lead to non-positive definite matrices in model estimation because of 'out-of-bounds' correlations, that is >1.0 in absolute value (Graham, 2009; Kline, 2005).

3.5.2. Single-Imputation approaches

Another general class of approaches are imputation based. The example of mean substitution given earlier is one method, where a value is imputed based on some characteristics of the dataset, in this case the mean value of the non-missing instances of the missing value. As mentioned, the implication of *imputing the mean* is that variance estimates reduce, as do correlations with other variables, and kurtosis being impacted by increasing the frequency of the central value (Vriens & Melton, 2002). Given the basis of SEM on variances and covariances, its use has been cautioned against (Brown, 1994; Arbuckle & Wothke, 1999). As the amount of missing data increases, error variance may also be underestimated because a single value is repeatedly entered as the missing items value (Vriens & Melton, 2002; Kline, 2005; Vriens & Sinharay 2006).

A second type of imputation, often used in model-based solutions is regression-based single-imputation. Byrne (2001) outlines this as a process in which regression equations, using the missing data as the dependent variable and present data as predictors, are used to generate values that would be expected. This approach uses the structure of the data to make the imputation. Ideally, the data structure should be related to the data loss mechanism, so that variables used covary with the missing data variable to some extent (Kline, 2005). Graham (2009) suggests that there may be loss of error variance from such a single imputation because imputed values are fit to lie on the regression line, reducing variability. Perhaps a positive of the approach is that it imputes values on a per-model basis, rather than for the dataset generally. So the imputing is based on an a priori idea of the relationship of a set of variables to each other, rather than on filling holes in the dataset without any theoretical framework.

Byrne, (2001); Graham, Cumsille & Elek-Fisk (2003); and others (Schumacker & Lomax 2004; Kline, 2005; Blunch, 2008) consider all of the foregoing, with some specific exceptions, to be generally unacceptable approaches to missing data management. Importantly these recommendations are not without caveats; deletion methods might be used in the initial stages of an exploratory factor analysis, for example (Graham, 2009). The remaining approaches are considered generally more acceptable as approaches; one of the key issues with them relates to whether a full dataset imputation is sought (e.g. using Multiple Imputation) or a specific model is being tested where missing data is present, where a Maximum Likelihood approach might be implemented.

3.5.3. Maximum Likelihood (ML) approaches

A more advanced model-based approach to data imputation, than the single-regression imputation method, is the use of what it termed an EM (Expectation/Minimisation) solution (Wiggins & Sacker, 2002). This involves two steps - Estimation (or Expectation) of missing values again by an initial regression based single-imputation. This is followed by a Minimisation step - where the whole dataset (now complete from imputed data) is ‘submitted for maximum likelihood estimation’ (Kline, 2005:55), which involves refining the best guess of the imputation values of parameters. As a result of the initial regression based single-imputation it suffers in terms of estimation of standard errors from the dataset, so hypothesis-testing procedures are cautioned against (Graham, 2009).

From within the field of structural equation modelling there have been a number of approaches, which have been tested in simulation studies (Peters & Enders, 2002; Wiggins & Sacker, 2002), with the conclusion that in general there is one approach that offers better performance in imputing missing data than the aforementioned treatments (Byrne, 2001; Wiggins & Sacker, 2002; Schumacker & Lomax, 2004:43). This approach is alternately called Maximum Likelihood (ML) imputation in LISREL (SEM software: Joreskog & Sorbom, 1999) or Full Information Maximum Likelihood (FIML) imputation for AMOS (hereafter ML). Wiggins and Sacker (2002) suggest that the FIML approach works well across 5-25% missing data, for both measurement and structural models. According to others (Arbuckle & Wothke, 1999; Schumacker & Lomax, 2004; Vriens & Sinharay 2006) the ML approach works in giving unbiased, or least-biased, estimates of all missing data treatments discussed, in both MCAR and MAR patterns of missingness. Additionally, even though the approach might lead to biased estimates in non-ignorable (MNAR) data imputation, the bias is least with this approach compared to the others (Muthen et al, 1987; Schafer, 1997; Byrne, 2001; Schumacker & Lomax, 2004). In a recent review of missing data handling, Graham (2009:573) proposed that Multiple Imputation and ML approaches should be used 'as a matter of course'. The final approach to data imputation considered focuses on the dataset, rather than the model being used, and imputes the missing data to the dataset, prior to the estimation of models of interest.

3.5.4. Multiple Imputation (MI) approaches.

As Graham (2009:556) states one critical component of multiple imputations is to '*restore the error variance lost from regression-based single imputation*'. Some simulation studies suggest it may give similar or better approximations with smaller sample sizes than the ML approach (Schafer & Graham, 2002). Given this imputation can be provided to the dataset prior to analyses of models, it offers the potential benefits of increasing power in the because of the inclusion of variables (in the imputation process) that might more accurately predict a missing value. Since these may not yet be included in any model being tested it follows an 'inclusive' variable strategy (Collins, Schafer & Kam, 2001; Schafer & Graham, 2002; Graham, 2009). The inclusion of variables related to the missing variable, but not necessarily included in the analysis, also potentially offers a chance to reduce estimation bias for non-ignorable (MNAR) missingness (Collins, Schafer & Kam 2001; Graham, 2009), a problem with all of the approaches discussed.

3.5.5. Missing data management in this study.

For each measure, an assessment of data missingness will be considered. However, because of the planned procedures involved in collecting data (see Chapter 4), it is proposed that missingness will be minimised; by checking incomplete questionnaire responses with participants, the only data that will be left missing will have a non-random cause. It is therefore appropriate to use a form of model based imputation, and depending on the extent of missingness this may be single imputation (for small amounts of missingness) or multiple imputation, for larger amounts, where it would be important to maintain the variability within the data. The extent of missingness is therefore a factor in selecting an approach; low levels of missingness generally mean more complex methods are less warranted (McCartney, Burchinal, & Bub, 2006).

3.6 Data Screening

The main issues with data quality, aside from missingness, relates to distribution; normality, skew, kurtosis and outliers. Non-normal distributions impact on variance and covariance of data, the core relationships in both Confirmatory Factor Analysis and Structural Equation Modelling (Schumacker & Lomax, 2004)

3.6.1. Univariate Normality

One approach to deciding on normality, presented by Field (2005), involves reviewing the critical ratio values associated with item univariate Skew and Kurtosis, and associated standard errors. Critical Ratios (*c.r.*) are derived by dividing skew and kurtosis values by their standard errors; these can be treated as *z*-scores (Arbuckle 2007). Given the assumption that a normal distribution will have a mean of 0 and a standard error of 1, the values can thus be compared for significant difference when converted to *z*-scores. With smaller sample sizes, it is appropriate to use a z-score cut-off at the $p< 0.01$ level (a value of 2.58 or greater; Field, 2005).

An alternative heuristic, based on absolute values of skew and kurtosis, suggests skew > 2.0 is moderately, and > 3.0 extremely, non-normal; for kurtosis > 7.0 is indicative of moderate non-normality, between 8.0 and 20.0, extreme (West, Finch & Curran, 1995:74; Osman et al 1997; Kline, 2005:50). Some authors suggest that kurtosis is the more significant concern with non-normality, in terms of impact on model estimation (West, Finch & Curran, 1995; Byrne, 2001). For kurtosis, Kline (2005:50) proposes a ‘conservative rule of thumb,’ that values > 10.0 in kurtosis might suggest a problem. Univariate normality does not preclude multivariate *non*-normality and so a consideration of multivariate normality is also warranted.

3.6.2. Multivariate Normality

Lack of multivariate normality (MVN) tends to inflate the computed chi-square (x^2) value (West, Finch & Curran, 1995). With higher x^2 values, there is greater difference between model-implied and actual data matrices, which are compared when model fit is being calculated (Schumacker & Lomax, 2004:66). Inflated x^2 could therefore lead to higher rates of model rejections (Finney & DiStefano, 2006:277) and reduced values for other measures of model acceptability. In conjunction with smaller sample sizes, non-normality can lead to solutions that fail to resolve, or converge (West, Finch & Curran, 1995; Byrne, 2001). On the other hand, departures from MVN can underestimate standard errors (West, Finch & Curran, 1995; Byrne, 2001). Smaller standard errors mean that regression paths, factor and error covariances can be statistically significant more often than they should be (West, Finch & Curran, 1995; Byrne, 2001).

Multivariate normality can be assessed using Mardia's coefficient of multivariate kurtosis in AMOS (Mardia, 1970; Mardia, 1974; Mardia 1980; Arbuckle, 2007). Mardia's (1980) coefficient of multivariate normality possesses an 'approximate standard normal distribution' (West, Finch & Curran, 1995:61), meaning a critical ratio of >1.96 represents a significant departure (at the .05 level). However, this figure is considered only a guide and the size of the score is not considered a good indication of the amount of departure from normality (Arbuckle, 2007).

3.6.3. Outliers

One final consideration is the impact of outliers, which might contribute to non-normality in the score distributions. Aside from ruling out data entry errors, options for treatment of outliers depend on whether they are univariate or multivariate outliers. The classification comes down to whether the scores on a single item (univariate) or set of items (multivariate outliers) would be expected to be from within the population under study. Therefore, detection might depend on the number of standard deviations from the mean, position outside the top/bottom quartile from the median, or a value that has a p value of < 0.001 on a x^2 distribution (Stevens, 2002; Kline, 2005). For univariate detection Stevens (2002) suggests that most values should be < 3.0 in absolute value in a *z*-score distribution, with an understanding that in a 'large' sample (>100) it might be expected that some would be > 3.0, just by chance.

Dropping data, which has been checked for recording or instrument errors, is not typically considered legitimate (Stevens, 2002:17). In part this is because deletion makes an assumption that the occurrence of this data is an anomaly, rather than of significance. Visual inspection will be the initial method for screening for univariate outliers in this study, followed by a consideration of whether action is warranted.

Multivariate outliers are more difficult to detect. These have extreme scores of one or more variables or a general atypical set of scores (Kline, 2005). One method used is judging the distance of a combined set of variables (e.g. values of multiple items on the Beck Depression Inventory) from a point in multidimensional space where the mean values on all 21 items (of the BDI) would intersect, and around which most values would be expected to collect (Fidell & Tabachnick, 2003); this is called the Mahalanobis Distance (d, with d^2 being the score presented; Fidell & Tabachnick, 2003:130; Kline 2005:51) and since, at least in large samples, the d^2 value has a Pearson x^2 distribution, a statistical value of p 0<.001 is considered a multivariate outlier.

3.7. Measurement Scale

While the 4 response options to each of the 21 Beck Depression Inventory items mean that the level of measurement is ordinal, a common assumption in such circumstances is that the underlying factor being measured is continuous, and therefore it is analysed as such (Quintana & Maxwell, 1999). Of all the measures selected for data collection, only the Beck Depression Inventory has such a restricted range of ordinal measurement (the MFQ has seven response categories), so this section considers depression throughout.

Bentler & Chou, (1987) in a review of the assumption of continuity of a scale, generally made in CFA and SEM, discuss how continuous measurement scale is relevant in two respects - the data gathered from any sample under study is not itself continuous; there are only as many pieces of the supposed continuum as there are individuals in any sample (Bentler & Chou, 1987:87). The second point relates to the ability to determine linear relationships between ordinally- and continuously- measured variables; they suggest that 4 or more categories were probably sufficient for a continuous treatment.

Kendall et al, (1987) make a relevant point regarding scoring on the Beck Depression Inventory (BDI) being representative of a latent continuum rather than discrete categories, despite its four-choice response categories. In effect, the underlying continuum of mood not only stretches towards more severely depressed (higher scorers), but also towards people who were *very not-depressed*; these people, the authors suggest, were those who tend to score 0 on all items. This is the reason that this latent variable is termed *Mood*, rather than Depression in this study. Thus, because the BDI is not considered diagnostic of categories of depression *per se*, the underlying distribution can be considered continuous; scores say that one person displays more depressive features than another. Higher scores indicate more sever depressive symptomatology than lower scores, not differences between, for example, Major Depressive Disorder and other Depression diagnoses.

The second issue in using categorical description of presumed continuous constructs, such as depression, relate to the number of possible response categories, the amount of skew and whether all variables were skewed in the same direction (Bentler & Chou, 1987; Byrne 2006). Byrne (2001) reviews the evidence on these issues and suggests that 'critical' problems with estimation of factor loadings and factor correlations arise when scales have 3 or less response options (the BDI has 4) and when skew values were greater than 1.0.

Alternative approaches to handling the ordinal nature of Beck Depression Inventory data have been used (Thombs et al., 2008), by explicitly modelling the Beck Depression Inventory items as ordinal data and estimating a polychoric correlation matrix. This approach has been discussed in the Structural Equation Modeling literature also (Byrne, 2001; Kline, 2005) and could be characterised as reflecting, in part, the debate on ideal sample size required for SEM. One view is that very large sample sizes are required (>5000; Bentler, 1994), in order to use such alternative estimation techniques for categorical data (Byrne, 2001:71) and an 'extremely strong' multivariate normality assumption goes with these alternative estimation methods (e.g. Asymptotic Distribution-fee estimators; Byrne, 2001), making it an approach, according to Byrne, that is more appropriate to simulated rather than real data. With normally distributed categorical variables and 4 or more response categories, that normal theory methods probably do not present significant costs, according to Bentler & Chou (1987).

3.8. The two-step approach to Structural Equation Modeling

The two-step approach to Structural Equation Modeling (SEM; James, Mulaik & Brett, 1982; Byrne, 2001:12; Schumacker & Lomax, 2004; 209) proposes two stages in the modeling process. First, the *Measurement Model* tests a series of confirmatory factor analytic (CFA) models for each of the latent variables; these latent variables are used in later structural model testing, step two. The CFA aims to confirm that the observed measures do contribute to the factorial structure of each of the proposed latent variables

(Schumacker & Lomax, 2004) and is therefore an assessment of convergent and discriminant validity (Brown, 2006; Kline, 2006).

The observed measures selected to represent the latent variables in this study have been presented in Chapter 2: *Development of Methods*. The set of processes for carrying out confirmatory factor analyses are the same as those for testing structural models and include model *Specification, Identification, Estimation, Fit, Modification* and reassessment of *Fit* (Bollen & Long, 1993; Schumacker & Lomax, 2004). While the process can be iterative (modification, re-estimation, further modification), models being tested at this stage are confirmatory, rather than exploratory, in that the expectations about factorial structure are specified *a priori*. Decisions about adapting a model for best statistical fit are guided, and limited, by theoretical constraints on model specification (Kline, 2006). For measurement models, additional considerations are based on published factorial studies; for structural models, the a priori models are informed by theory. Finally, outputs from the statistical analysis support model acceptance or rejection at the statistical level (Schumacker & Lomax, 2004; Blunch 2008). These latter statistical guides include; parameter estimates, inter-factor correlation, r^2 (squared multiple correlation) values, Standardized Residual Matrix assessment and various indices of model fit (Joreskog & Sorbom, 1989; Schumacker & Lomax, 2004; Yuan, 2005; Steiger, 2007). The stages in model development, testing and modification are now discussed, followed by a summary of the statistical indices of model fit used in the study.

3.8.1. Model Specification.

Model specification is the description of parameters that are of interest, and proposed to be true in the data, that the investigator wishes to confirm or investigate (Kline, 2006). The specified model is based on a theoretically guided expectation of the relationships within the data, and this is tested to establish the extent to which these relationships fit with the theory implied relationships (Schumacker & Lomax, 2004). Effectively, this specification sets out the model-implied variance/covariance matrix and this is compared to that given by the real data at the testing stage (Schumacker & Lomax, 2004). Prior to estimating the model, it has to be checked that a solution can be derived; this process is called Model Identification.

3.8.2. Model Identification.

The question in model identification is whether there is enough data to estimate the parameters of interest (Blunch, 2008). A common example (Schumacker & Lomax, 2004; Kline, 2005; Blunch, 2008) provided to explain identification is to imagine solving an equation e.g. x + y =10, for values of x and y. If x and y are considered parameters, there are more parameters than data (10 being the data). Values for x and y cannot be derived and so the model is not solvable; it is indeterminate in that an infinite set of values for x and y could be found (Schumacker & Lomax, 2004). This model would therefore be considered *underidentified*.

An additional parameter would help with identification at this point, so if, in addition to x + y = 10 we also include x - y = 2, the model now is considered *just-identified* in that its parameters are solvable (Schumacker & Lomax, 2004) Here the solution of x=6, y=4 is unique, and would indicate a perfect fit of the model (x + y; x - y) to the data (10). A third data value x + 2y = 16, leads to an *over-identified* model, for which a unique set of values for x and y are impossible to derive. Instead, using a fitting function a unique set of values might be achievable, though they may not exactly reproduce observed data (Kline, 2005). This is the job of model estimation, using a minimisation function (Arbuckle, 2007). With more and more parameters, and assuming there is sufficient data, the degrees of freedom within the model will increase, allowing for more refined fitting of the model (Blunch, 2008). Thus, higher degrees of freedom typically help with model fitting (Blunch, 2008); at the same time with more parameters to be estimated less exact values can be determined (Kline, 2005). Therefore the relationship between degrees of freedom and model fit is one that needs to be considered in accepting model fit (Raykov & Marcoulides, 2006). The SEM software typically provides estimation of identification.

A second aspect of identification for CFA is the setting of a measurement scale for each latent variable being modelled (Schumacker & Lomax, 2004). The requirement to do this is so that a reference metric can be set from which to estimate factor loadings for the latent variables, scaling the factor (Kline, 2005) it is achieved by setting one loading for each factor to equal 1.00. This is automatically completed in some SEM software, but can be changed if there is an indicator (observed variable), which has more reliable scores than another.

3.8.3. Model Estimation

In this study, all Confirmatory Factor Analyses, and subsequent structural models are tested using Generalised Least Squares (GLS) estimation, using the AMOS package (Arbuckle 2007). The GLS estimation approach is selected above the alternative Maximum Likelihood Estimation (MLE) approach because it operates ‘under less stringent multivariate normality assumption’ (Kelloway, 1998; Schumacker & Lomax, 2004:69; Blunch, 2008;). Maximum Likelihood Estimation process may produce more inflated x^2 values, compared to Generalised Least Squares, when dealing with data typical of behavioural research (West Finch & Curran, 1995).

3.8.4. Assessment of Model Fit.

The concept of *Model Fit* is based on the relationship of the model-implied, compared to data-implied, covariance matrix being different (Schumacker & Lomax, 2004). This is considered from two perspectives The first assessment, of global fit, tests the fit of the entire model and second, an examination of the fit of individual parameters of the model, includes an assessment of the parameter estimates’ significance and sign, the standardised residual covariance matrix and other features (Kline, 2005; Schumacker & Lomax, 2004; Byrne, 2001:74). These approaches to the assessment of fit will be discussed now, bearing in mind that the context for these judgements must be a theoretically plausible model in the first instance.

3.8.5. Fit Indices

A general heuristic is that a range of model fit indices should be considered, and reported. These are discussed and their ideal value ranges outlined, bearing in mind that *‘the array of writings on overall tests of model fit is extensive and not all together consistent’* (Maruyama, 1998:239).

3.8.5.1.Chi-square value, p-value and Chi-square to degrees of freedom ratio (x2 /df)

The chi-square (x^2) and its significance value (p) are considered important and generally required indices in terms of estimating global model fit. They are useful because they provide a statistical test of the model (Schumacker & Lomax, 2004).

A key component of establishing model fit is the comparison of two matrices (data-implied and model-implied), using a x^2 comparison (Schumacker & Lomax 2004). An acceptable model will have a data covariance structure similar to the model being tested; the data will fit the model leaving a residual of the proposed model covariance matrix minus the data covariance matrix of zero, or not significantly different from zero (Byrne, 2001). Therefore, the difference between model-implied and data-implied covariance matrices should *not* be significantly different from zero if the model is an acceptable explanation for the data, that is, the p value of the x^2 test should be >0.05 (Byrne, 2001). Thus, in a well fitting model, the p value of x^2 behaves as a 'badness of fit' index (Schumacker & Lomax, 2004:82, 100; Hu & Bentler, 1995; Kline, 2005).

The x^2 test is the most basic fit statistic and while it is typically reported it is not without some limitations, notable in respect of distortions caused by sample size (Schumacker & Lomax, 2004; Kline, 2005). Samples above 200 and below 100 may cause x^2 estimations of significance and non-significance respectively (Schumacker & Lomax, 2004). One additional consideration with the x^2 test is the null hypothesis it tests against is probably too strict given the types of models being tested; the idea that any model fully accounts for performance in the population as a whole is probably unrealistic, for example, in the social sciences (Bollen & Long, 1993; Kline, 2005).

The statistic is traditionally reported despite the proposed limitations and it is also used in some instances to derive other measures of model fit. Use of the x^2 to degrees of freedom (df) ratio (in AMOS this is denoted as, CMIN/df) is suggested as a potentially useful index of fit (Marsh, Hau & Grayson, 2005). Values for CMIN/df should be < 2.0 (Kline 2005), or 'near 1.0' (Blunch, 2008:113).

3.8.5.2. GFI: Goodness of Fit Index & CFI: Comparative Fit Index.

Many of the measures of model fit are strongly related to the degrees of freedom (df) within the model, as well as x^2 (Hox, 2002). Thus, measures of fit based in part on df, such as GFI and CFI (Steiger, 1990), are impacted by the df number.

CFI values tend to decline as numbers of variables are added to correctly specified models (Kenny & McCoach, 2003). Expected values for good fit are > 0.90 for both GFI and CFI (Marcoulides & Hershberger, 1997; Schumacker & Lomax, 2004; Kline, 2005; Hu & Bentler, 1995). Both indices indicate how well covariance and variances in the sample matrix are explained by the proposed model (Byrne, 2001), or how much better the proposed model fits the data matrix, than no model for GFI (Mueller, 1996; Joreskog & Sorbom, 1998) and a null model for CFI.

A null, or independence model is one where no covariance among observed variables are assumed, that is, unrelated variables (Kline, 2005; Joreskog & Sorbom, 1998) The CFI was developed in an attempt to minimise the impact of sample size (Bentler, 1990) and also to compare models not against no model, but the independence or null model (Byrne, 2001). One benefit of this index of comparative fit is that it does not assume zero error in the approximation, i.e. perfect fit, in the baseline model (Kline, 2005).

3.8.5.3.Root Mean Square Error of Approximation, 90% confidence intervals & p value

One of the proposed benefits of this measure (RMSEA; Steiger & Lind, 1980 cited in: Steiger, 1990) is that it is not based on a hypothesis that the proposed model is 100% correct, which is largely untenable (Brown & Cudeck, 1993; Blunch, 2008). Instead of focusing on fit, it focuses on the level of discrepancy, measured by residuals. In this sense it indicates the 'degree of lack of fit', taking the number of parameters into account (Brown & Cudeck 1993:137). Its benefit is in being the main other measure, aside form the x^2 test, which has a statistical test of its value, and offers both a point estimate value, and a confidence interval (Kelloway, 1998). What the index aims to measure is given by Brown & Cudeck (1993:137) as *'How well would the model, with unknown but optimally chosen parameter values fit the population covariance matrix if it were available.'* The range of acceptable values for RMSEA is debated in the literature, with acceptable values suggested to be <0.10 (MacCallum, Browne & Sugawara 1996), <0.08 or <0.05 (Byrne 2001:85). Steiger (2000:161) suggests that cut-off 'should not be taken too seriously'.

Of more interest are the 90% confidence intervals (90% CI; Steiger, 1990; Steiger, 2000) around the value, especially when comparing models, and the p-value of the RMSEA (ideally $p > 0.50$; Byrne 2001). All these values will also therefore be reported. One cautionary note relates to model complexity, and sample size, and their influence on the confidence intervals (MacCallum, Browne & Sugawara 1996; Byrne, 2001). If sample size is small and model complexity high (many parameters) then wide confidence intervals are likely. This may be a concern when CFA modeling the Beck Depression Inventory, which as 21 observed items.

3.8.5.4. SRMR: Standardised Root Mean Square Residual

A final index of model fit, recommended by Kline, (2005), is the SRMR (Standardised Root Mean Square Residual), defined as *'the average discrepancy between the correlations observed in the input matrix and the correlations predicted by the model'* (Brown, 2006:82). Kline (2005) states that values of < 0.10 are favourable, and Schumacker & Lomax, (2004) and Byrne (2006) suggest values < 0.05. One proposed benefit of SRMR is that it is less sensitive to the effect of sample size than the non-standardised RMR (Root Mean Square Residual; Marsh, Balla & McDonald 1988; Marsh, Hau & Grayson, 2005). Others suggest that this insensitivity to sample size is true with large samples and misspecified models, but not true with small samples where models are generally 'acceptable; but might have some mis-specification' (Marsh, Hau & Wen, 2004:335). The SRMR will be used as an index of fit in the following model testing procedures, but consideration needs to be given to some CFA models likely being generally acceptable (they are in part, based on previous studies' findings) with the possibility of some mis-specification (different population) in conjunction with small sample (n<150; Marsh, Hau & Wen, 2004). In general it is advised that a review of the 'raw' standardised residuals themselves is also carried out (Grover & Vriens, 2006). This is achieved by reviewing the *Standardised Residual Covariance Matrix* which will be reported with each model tested.

3.8.5.5. Other indicators of model fit.

In addition to the x^2 fit test, and the fit indices discussed, interpreting individual parameters for statistical significance, size and expected direction is warranted. For CFA models, interfactor correlation is also an important assessment. Additional information about model fit can be gleaned from inspection of the Standardised Residual Covariance Matrix and the r^2 (coefficient of determination; Steel & Torrie, 1960) values. Using these measures, fit is assessed *by components* (Bollen & Long, 1993:6).

The Standardized Residual Covariance Matrix can be used to look at discrepancies between data-implied and model-implied item covariances (Schumacker & Lomax, 2004). High values (> 2.58; Joreskog & Sorbom 1998) can indicate problems. It is therefore possible to inspect for item covariances generating high residual covariances with some specificity in terms of which items might underlie the poor fit. Arbuckle, (2007) recommends that the majority of these values should be < 2.

In a Confirmatory Factor Analysis, indicators of r^2 < 0.20 should indicate possibilities of dropping those items (Schumacker & Lomax, 2004) that is, 20% or less of an items variance is accounted for in the factor loading, or conversely 80% is unexplained. A final reason that model fit might be questioned is based on the individual parameter estimates not being significant, when they were expected to be (Maruyama, 1998). These parameters might include expected correlations or specific causal paths in structural models.

Table 3.3 summarises the range of assessments to be used in appraising both Confirmatory Factor and Structural Equation models, and ideal values for the specified fit indices are summarised in table 3.4.

Table 3.3: Fit indices and other assessments of model acceptability. x^2 = chi Square value; x^2 p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual.

Specific indices of model fit	**Other Assessments**
x^2 & p-value	Standardized Residual Matrix values < 2.80
CMIN/df	Parameter sign & significance
GFI	Squared Multiple correlation *(r^2)* values > 0.20
CFI	
RMSEA: 90% CI & p-value	
SRMR	

Table 3.4: Summary of recommended values for model fit indices. x^2 = Chi Square value; x^2p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual.

x^2 p-value	**CMIN/df**	**GFI**	**CFI**	**RMSEA**	**SRMR**
p > 0.05	< 2.0	> 0.90	> 0.90	90% CI upper limit of 0.08; pclose > 0.50	< 0.10

3.8.6. Model Modification

A number of elements of the statistical outputs can be used to identify modifications to a model. These include modification indices, used to detect relationships that may optimally reduce the x^2 value, thereby potentially improving model fit. Modification indices are produced by AMOS and suggest which parameter changes would give the greatest change in the x^2 value of the model, potentially improving fit. The modification indices are based purely on statistical grounds, and bear no necessary relationship to the theory underpinning the model itself (Schumacker & Lomax, 2004). For this reason many suggested modifications are not theoretically plausible. Typically, modification indices might indicate error term covariances, factor cross loadings or incorrect item-factor loadings, but until tested there is often no indication of the values of parameter estimates in the newly modified model (Schumacker & Lomax, 2004). Kelloway (1998:38) discusses modification indices in terms of 'theory building and theory trimming', using backward elimination and forward entry stepwise regression. The implications of the analogy are the potential for inflated error and data-dredging approaches to theory building (Bolker, 2008), that is, a non-confirmatory approach being pursued (MacCallum, 1986).

In a strict confirmatory process, modifications are not considered appropriate (MacCallum, Roznowski & Necowitz, 1992), but in reality some post hoc model modifications often do take place. Here, it is recognised that where a number of factor models have been presented in the literature, dimensionality may be different in different samples, so a conservative approach to modification is carried out, bearing in mind the measures of fit may be impacted by the size of the sample and the distribution of the data, and not a poor model (Maruyama, 1998).

3.9. The Statistical Assessment of Mediation

Figure 3.4 presents a generalised mediation model. An initial assessment for the potential for mediation is based on Iacobucci (2008), Baron & Kenny (1986) and MacKinnon (2008) who broadly indicate that there should at least be a significant relationship between predictor and mediator and/or mediator and criterion variable. The prerequisites for assessing mediation have been debated and are discussed from a strict perspective by Barron & Kenny (1986), or with fewer requirements by MacKinnon (2008) and Iacobucci (2008), based on how the mediation is tested and interpreted. One difference from Baron & Kenny is the proposal that there need not be a significant direct effect of a predictor (exogenous variable) on a criterion (endogenous variable) in order of mediation testing to be carried out. Such occurrences are proposed to reflect a suppressing mediator or one in which with its effect removed (partialed), the true relationship can be uncovered (MacKinnon, 2008). A mediational process is tested for by use of the Sobel *z*-test (Sobel, 1982; Iacobucci, 2008). How the Sobel *z*-test is calculated and interpreted is discussed below.

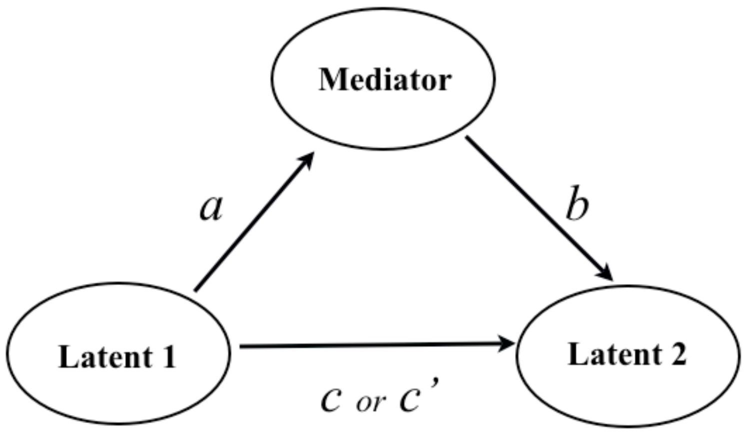


Figure 3.4 Summary of mediation testing *c* = the direct effect of *Latent 1* on Latent 2; c' = the effect of *Latent 1* on Latent 2, when the mediated effect of *Mediator latent*. *a*b* should be equal to *c minus c'*. If the difference between *c* and *c'* is statistically different from zero, then mediation can be suggested

Statistically, this is tested as an indication that *c* (the direct effect of *Latent 1* on Latent 2) is no different to c' the effect of *Latent 1* on Latent 2, when the mediated effect of *Mediator latent* is accounted for, *a*b* being equal to *c minus c'*. If the difference between *c* and *c'* is statistically different from zero, then mediation can be suggested (Iacobucci, 2008; MacKinnon, 2008). The Sobel *z* score is calculated using the unstandardised regression weights and standardised errors of paths *a* and *b* as outlined in figure 3.5. The values for parameters are populated from the SEM modelling output. This approach provides for more accuracy, compared to a series of individual regression equations that would be required in the presence of multiple indicators of latent variables (Iacobucci, 2008). The test formula and calculation tool is given by Preacher & Leonardelli (2006):

Sobel test: $z\text{-value} = a*b/\sqrt{(b^2*s_a^2 + a^2*s_b^2)}$ *formula.1*

Where a is the unstandardised regression coefficient from the exogenous variable to the mediator, b between mediator and endogenous variable. s_a and s_b are the standard errors for the their respective paths, a and b The p-values are drawn from the normal distribution with assumption of a two-tailed z-test of the hypothesis that the mediated effect equals zero in the population, with +/- 1.96 are the critical values. (Preacher & Leonardelli, 2006). An applied adjustment is proposed by MacKinnon et al, (2002) and MacKinnon, Lockwood & Williams (2004) who suggest mediation in samples of n=100 for the Sobel z test exists where the absolute critical value is > 0.97 (equivalent to two-tailed $p < 0.05$). This adjustment in interpreting the critical values provided by the Sobel z test, is based on findings that the derivation of the Sobel tests by dividing regression coefficients by their standard error provides low statistical power and assumes a normal distribution in the indirect effect (MacKinnon, Lockwood & Williams, 2004; MacKinnon, 2008). The cut-off of ± 0.97 will therefore be used in assessing mediation.

3.10. Summary

In the development of statistical methods three main issues were considered. First, defining what cognitive impairment and metamemory accuracy would be for the study. Second, considering adequate sample size, the management of missing data and the screening of data to address the assumptions of the proposed statistical method. Finally focusing on the two-step method of confirmatory factor analysis and structural modeling, addressing how acceptable fit is assessed.

Chapter 4: Methods

4.1 Introduction

This study used a cross-sectional design to investigate the factors that contributed to metamemory in a sample of community-dwelling people with Multiple Sclerosis. The purpose of the design was to investigate the relationships between measures of cognitive function and mood, and a set of measures of accuracy in memory judgments. This chapter will summarise the participants, the selected materials and the procedures used for capturing data for the study.

4.2. Specification of Sample Characteristics

An important consideration for the study is of the sample characteristics. Here a brief consideration of the requirements in terms of sample characteristics is presented to underpin the generation of inclusion and exclusion criteria.

In respect of the focus on subjective appraisal of memory function, two considerations are important; first that participants have a relatively stable experience of their abilities, as far as disease progression allows. Secondly recent or radical changes in ability have not occurred. To this end, people who are currently undergoing a relapse will be excluded from taking part. In order to minimise this type of participant recruitment was not planned using service providers (e.g. NHS, general practitioner), which might, in the main, include people who have had recent changes in function.

Given the range of difficulties with sensory and motor performance, which might make participation difficult, and increase measurement error, tasks were specifically chosen to widen accessibility. More severe limitations, which would make participation too difficulty, and measurement unreliable, were therefore appropriate as exclusion criteria.

A final consideration was of co-morbidities, which might confound interpretation of the results of the study. The presence of other neurological diagnoses would make conclusions about the impact of MS on memory appraisal difficult. However, more difficulty arises with mood disorder. The relationships of euphoric states with brain damage, hypomanic states with steroid treatment, increased incidences of bipolar disorder among people with MS (Minden & Schiffer, 1991) and the poorly defined genesis of depression in MS, make clear exclusion criteria difficult to define. A psychiatric diagnosis prior to diagnosis of MS was therefore considered an appropriate exclusion, but subsequent diagnosis, especially of mood disorder, was not.

A second consideration is accessing a range of participants, both those who tend towards ‘involvement’ and those who do not, or do so to a lesser extent. Recruitment avenues were therefore considered carefully. Advertisement through support organisations were more likely to recruit those likely to be support seekers, reflecting the limitations of the expert patient programmes discussed earlier, whereas recruitment through magazines might better access those who, although members of organisations, might be less actively involved in support seeking. Relatedly was a consideration of how to recruit participants so as not to bias the study itself - if a screening process for cognitive impairment was carried out, this would likely exclude participants for whom affect related biases might be extant. The author was mindful too of the inconclusive literature about how related memory and metamemory impairments are in neurological populations. As a result the advertising required that it seek out a ‘typical’ community dwelling population as far as possible. There is no sampling frame for people with MS in the UK to guide selection (O’Hara, De Souza & Ide, 2000) of representative samples. In seeking people who have concerns about their memory as a general criterion, it was hoped that subjective appraisal would guide interest and therefore capture the *concerned, but without memory deficit*, as well as those *concerned, and with memory deficit*.

The sample was a convenience sample of people with Multiple Sclerosis who were living in the community. In line with sample size recommendations for the statistical methods, a minimum of 100 was sought (Schumacker & Lomax, 2004). Participants were recruited through advertising directed to a number of support organisations, publications, websites and the MS Therapy Centre network in London and surrounding counties. Participants were given a £15 gift token for taking part.

For inclusion in the study, participants were required to have a confirmed diagnosis of MS, be over 18 years old, and have no history of other neurological diagnoses. In addition, there were required to be community dwelling, have sufficient visual, writing and speech abilities to engage in tasks, sufficient ability to read and understand words, sentences and instructions in English (tests are normed for English speakers). A history of psychiatric diagnosis prior to the onset of MS, and a history of alcohol or substance misuse were exclusion criteria. Advertising sought people who were interested in a study abut how well they knew their memory.

4.3 Materials

In Chapter 2 the selection of measurement tools was discussed. Here, relevant administration and scoring details are outlined for each of the *Memory, Information Processing, Executive Function, Mood* and *Metamemory* instruments.

4.3.1 Measures of Memory

Memory tests were interleaved with metamemory judgments. For clarity, the memory component of each is described.

Auditory Verbal Learning Test (Lezak, 1983).

This task required the learning and recall of two 15-word lists. The 15-word stimuli have been adapted into English (Lezak, 1983; Lezak, Howieson and Loring 2004), and consist of common high frequency nouns. The 15-word list (List A) was presented orally, with a one-second interval between each word. The participant received 5 presentations of word list A in total (Trials I-V), with a test of recall of the full list after each presentation. No feedback is given on performance between trials. After the 5 trials on List A, a second list of 15 new words is introduced for immediate recall (List B, Trial VI). After a 20-minute delay, a recall trial (Trial VII) is carried out on the original list of 15 words (List A).

Using the scoring guidelines of Lezak, Howieson and Loring (2004:422-423) number of words correctly recalled after 20 minutes was the main scoring index for latent variable modelling.

The Sentence Memory Test (Nelson 1984; Shimamura & Squire, 1986)

In this task, both a recall and recognition test was provided. Participants were required to remember the final word of 24 sentences, each sequentially presented centred on a landscape A4 page in 30-point bold Century Gothic font. Verbal stating of each sentence was required from the participant, and the examiner provided an addition 2-second interval before turning to the next sentence. After a delay of 60 minutes, filled with a break and other cognitive tasks, the sentences were re-presented in a different order each having the final word missing. Participants were asked to provide their answer for the missing sentence completion word. The recall score was the number of correctly completed sentences.

To test recognition performance, participants were then presented with those sentences that were incorrect, or to which *don't know* answers were given. They were given a choice of eight plausible completion words for each sentence and asked to select the correct answer. The number of correct selections was their recognition score.

Participant instructions for the task were as follows:

> *'In this test I am going to show you twenty-four sentences. Your task is to read the sentence out loud, and try to remember the last word of each sentence. Your ability to remember the last word of each sentence will be tested later on. This is therefore a test of your memory.'*
>
> *'I will give you a practice sentence to start with, and I will explain the process. Then we will go through the whole set of twenty-four sentences'*
>
> *'In about an hour, I will be asking you to try and remember the last word that went with each sentence. This will be tested by showing you each sentence again, this time with the last word missing. These will be in a different order to the way you saw them the first time'*
>
> *'Now, let's do an example'*

Participants were then shown the sample sentence 'She put the flowers in the VASE'

> *'Here is a sentence with the key word for you to remember in large type. I want you to read this out and try to remember this word, because when you see the sentence again it will be missing'*

Participants were shown the sample sentence again, with the final word missing 'She put the flowers in the _________'

> '*If you cannot remember the word, or get it wrong, I will give you a set of possible answers, only one of which is correct, to see if you can pick out the correct one, like this:'*

Participants were given the incomplete sentence again, with 8 possible answers, and their attention was drawn to the fact that all of the choices are plausible answers - Garden, Water, Kitchen, Coffin, Soil, Vase, Bin, Window.

'You will notice from the eight choices that only one is correct, Vase, but that all of them could fit and make sense of the sentence. This is important because it means the answer will not be obvious, unless you have some memory of it'
'Do you understand what I am going to ask you to do?'

This task was also used to generate Retrospective Confidence Judgments and Feeling of Knowing judgments, the exact procedure for which is outlined later in this chapter. The full set of test materials, including recognition foils used in the Feeling of Knowing test, are given in Appendix A.

4.3.2. Measures of Information Processing.

Symbol Digit Modalities Test (SDMT; Smith, 1982).
Participants were presented with rows of meaningless symbols on a single 11" x 8 ½ " page. The task was to convert these to their appropriate number, by referring to a symbol/digit key, presented at the top of the page. Here, the oral version was administered; participants stated the number associated with each symbol instead of writing it down. A practice set of 10 symbol to number conversions was given and participants were notified of any mistakes they made. Scoring was the number of correct symbol to digit conversions made in 90 seconds, not including those carried out in the practice set. Higher scores indicate faster processing speed.

Digit Span (Forward) task (Wechsler, 1997b)
The Digit Span forwards task required participants to immediately repeat increasingly longer lists of numbers, presented verbally by the examiner. Digit lists are reported back in the same order as they are presented. Digit spans start at 2 numbers and increase in sets of two, to 9 numbers. The task is stopped if a participant got two number strings of the same length incorrect. Scoring was the number of correctly recalled digit sequences. Maximum score is 16, with higher scores indicating better performance.

Letter-Number Sequencing (Wechsler, 1997b)
The Letter-Number Sequencing (LNS) task is taken from the Wechsler Adult Intelligence Scale 3rd Edition (WAIS-III, Wechsler, 1997a) and required participants to reorder a series of mixed letters and numbers, which increase in length by one item on each trial.

Participants were verbally presented letter number mixes and are asked to reorder them (mentally) and report them back. Numbers have to be verbally reported back first, in increasing order of size, followed by letters in alphabetical order. Scoring was based on the number of sequences correctly reorganised. Higher scores on the LNS task represent better performance. The maximum score possible is 21.

4.3.3. Measures of Executive Function

Brixton Spatial Anticipation Test (Burgess & Shallice 1997).

The Brixton Spatial Anticipation test requires participants to extract a rule from looking at the changing position of a blue circle among nine additional white circles, set in an array of two rows of five circles. As a 56-page stimulus book is worked through, the location of the filled circle in the 2 x 5 matrix changes. Each circle location is numbered 1-10. The participant looks at each page and indicates where they think the blue circle will be positioned on the next page, based on any patterns that they have detected in the movement of the blue circle. In accordance with the instructions provided in the test manual, participants were told that there are patterns as to the way the circle moves around, but that they have to establish what those patterns are. Additionally, they were told that an established pattern may change without notice, and, when this happens, they were required to figure out the new pattern governing the movement of the blue circle. Participants indicated their prediction of the upcoming location of the blue circle by pointing to the location and saying the number of the predicted location on the next page of the stimulus book. The researcher then turned the page.

Scoring was the total number of errors made in estimating the patterns of location of the target circle. An error is an incorrect prediction of the location of the circle, except in those situations where the pattern changes; if the predicted location matches the location that would have been correct if the 'old' pattern had been maintained, then that is a correct response. The first prediction is correct regardless of location.

Hayling Sentence Completion Test (Burgess & Shallice 1997).

This test comprised two parts. For Section 1, the test required participants to verbally complete a set of 15 verbally presented sentences with any relevant final word. Participant instruction and scoring was in line with those outlined in the test manual (Burgess &

Shallice 1997). Two sample sentences were given as practice items – e.g. "The rich child attended a private..." The participant was asked to offer a final word that completes this sentence in a sensible way (e.g. "School"). Time taken to respond is recorded. This section gives a response speed score. Once complete, a new set of 15 sentences (Section 2: Inhibition) is provided, now with the requirement to complete sentences with an unrelated word. Again, two examples were carried out before the test begins, e.g. "Most sharks attack very close to..."

Scoring on Section 2 was based on classification of the required inappropriate completion word according to its level of unreleatedness to the sense of the sentence (Errors Section 2), as well as how long it took to produce a response (Response Time Section 2). Errors were classified based on their extent of unreleatedness to the obvious sensible completion. Category A errors are either sensible completions (‘that would be unsurprising if encountered in a story or other narrative’; Burgess & Shallice, 1997:9) or responses which are ‘obviously rude or inappropriate’ (Burgess & Shallice, 1997:9). Category B errors include answers which have strong semantic relatedness to the sentence, are contrary words, including opposites, or words which connect with the last word of the given sentence (e.g. "Most sharks attack very close to... *the edge*"). Examples of acceptable answers for ‘Most sharks attack very close to...’ might be *potato*, or *yellow*.

In all, four measures were yielded, proposed to reflect aspects of executive function; response initiation (sum of response times in section 1), time taken to respond section 2, and error score in section 2. All three of these scores can be combined to generate the fourth measure - an overall score of performance. Here, the *Overall Score* will be used as an indicator of executive function for between-participant comparisons. This score reflects the trade-off that is often made between speed of response and accuracy of response in Section 2.

For structural modelling, three measures will be used from this task. Hayling Time will be calculated as response speed in the Inhibition condition *minus* response speed in the easy condition, to avoid confounding executive difficulty (indicated by speed) with information processing ability. The other two measures will be numbers of A-type and B-type errors.

For both executive function measures, higher scores indicate more impairment. In assessing the measurement and structural models, these scores are reversed so that higher scores indicate ability. This reversal was carried out to maintain coherence with higher scores indicating better performance in other cognitive domains.

4.3.4. Measure of Mood

The Beck Depression Inventory II (Beck, Steer & Brown 1996)

The Beck Depression Inventory Version II (BDI-II) was used to measure self-reported symptoms of depression. It consists of a set of 21 statements against which participants indicated their experience on a 0-3 scoring scale. Higher scores indicate greater levels of self-reported depression.

Participants were asked to complete each item considering a time frame of '*the past two weeks, including today*' in line with the current DSM-IV criteria for major depression (Beck, Steer and Brown, 1996:7; American Psychiatric Association, 2000). Participants were given the inventory to fill out in the presence of the researcher. If participants had difficulty choosing between two options, they were asked to select the higher number score. For those participants who had difficulty with either reading or writing (e.g. with pen control), the researcher read each question. Scoring was the sum of scores across all 21 items, with higher scores indicating more self-reported symptoms of depression. Total score on the scale will be used as the index of current mood for between-participant analyses. Scoring on each of the 21 items will be used in confirmatory factor analysis.

4.3.5 Measures of Metamemory

Two categories of measures were used to index Metamemory. First, three task-based memory judgements – Judgment of Learning (JoL) based on delayed recall of the 15-word Auditory Verbal Learning Test, a Retrospective Confidence Judgment (RCJ) and a Feeling of Knowing (FoK), both derived from the Sentence Memory Task. The second category was memory self-efficacy, based on the Memory Functioning Questionnaire (MFQ)

4.3.5.1. Judgment of Learning

The Auditory Verbal Learning Test (AVLT) has been outlined. The calculation of the Judgment of learning was based on *predicted versus actual* recall on the delayed recall trial of this test. Relatedly, in order to derive an indication of sensitivity to repeated trials on this memory task, the underconfidence-with-practice effect was investigated. Prior to presentation of each word list of the AVLT, participants were asked to estimate how many of the 15-word list they thought they would correctly recall. Underconfidence-with-practice was established by assessing the change in prediction discrepancy across all seven trials. The measure of Judgment of Learning accuracy was the difference between predicted and actual scoring on the delayed recall trial (Trial VII)

4.3.5.2. Retrospective Confidence Judgment.

Recalled items from the Sentence Memory Test were used for this judgment. When participants were asked to recall the sentence completing word for each of 24 sentences, they were also asked to make a judgment of their confidence in the retrieved answer, using a rating scale of 4, 3, 2 or 1 to indicate their level of confidence, where 4 indicated full confidence and 1 indicating a pure guess, no confidence. An A4 card, containing the 4 rating definitions, was given to each participant during the recall trial so that they could indicate their confidence level. Figure 4.1 shows the rating card given to participants so that they could indicate their rating of confidence.

How *confident* are you that you are correct?

4 = High Confidence/ I am correct

3 = Medium Confidence / Probably correct

2 = Low Confidence / Unlikely to be correct

1 = No Confidence / Guess

Figure 4.1 Retrospective Confidence rating card. This was given to participants during recall attempt on 24 Sentence Memory Test items.

Relative accuracy was calculated using the Goodman-Kruskal correlation based on the cross-tabulation method discussed in Chapter 3, and is reported as a non-parametric correlation ranging from -1.0 to +1.0, with 0 indication no-correlation.

Absolute accuracy was calculated by taking the proportion (in percent) of correctly recalled items given a high confidence rating (accuracy), and the proportion of correctly recalled items in the Low confidence condition (inaccuracy).

4.3.5.3. Feeling of Knowing Judgment.

Non-recalled items from the Sentence Memory Test were used for this judgment. Prior to completing a recognition trial for unrecalled items, participants were given a 4-point rating scale on a landscape A4 page. Participants were then reminded about the sample recognition trial they completed at the beginning of the sentence memory test, and the seven alternative, and plausible, options they would see when making the recognition decision. The sorting card given is presented in figure 4.2.

What are the chances you will know the correct answer, when you see it?

4 = High Chance / Will recognise it
3 = Medium Chance / Probably will recognise it
2 = Low Chance / Probably will not recognise it
1 = No Chance / Guess

Figure 4.2 Feeling of Knowing rating card. This was given to participants prior to recognition attempt on any unrecalled Sentence Memory Test items.

This scale, with 4 indicating a High Feeling of Knowing, and 1 indicating no/very low Feeling of Knowing was kept in view while all unrecalled sentences, still missing their final word, were given to the participant to sort according to the rating scale. Once sorted by ratings, participants were then asked to sort within ratings in line with ranking procedure outlined by Shimamura & Squire (1986) and Nelson & Narens, (1990) that is; the highest 4 down to the lowest 4, highest to lowest 3, 2 and 1, until a full ranking of sentences was achieved, from highest FoK to lowest FoK. If, during the task, a participant decided to move items from one rating to another, this was allowed.

After the rating and ranking procedure was complete, the recognition trial was carried out. For each item, the unrecalled sentence was presented with the final word missing, this time with the eight possible completion words printed below. Participants were asked to state and point their selected word. The set of 8 choice words were derived from the range of words generated by the UK sample used by Arcuri, et al., (2001), but where 7 alternative completions from that study were not available, additional semantically related words were added.

Relative accuracy was calculated using the Goodman-Kruskal correlation for the *ranked* list, and is reported as a non-parametric correlation ranging from -1.0 to +1.0, with 0 indication no-correlation. Absolute accuracy was calculated, using the 1-4 *rating* list by taking the proportion (in percent) of correctly recognised items given a High FoK rating (accuracy), and the proportion of recognised items in the Low FoK category (inaccuracy). Figure 4.3 summarises the how the Sentence Memory Test was used to derive the RCJ and FoK measures.

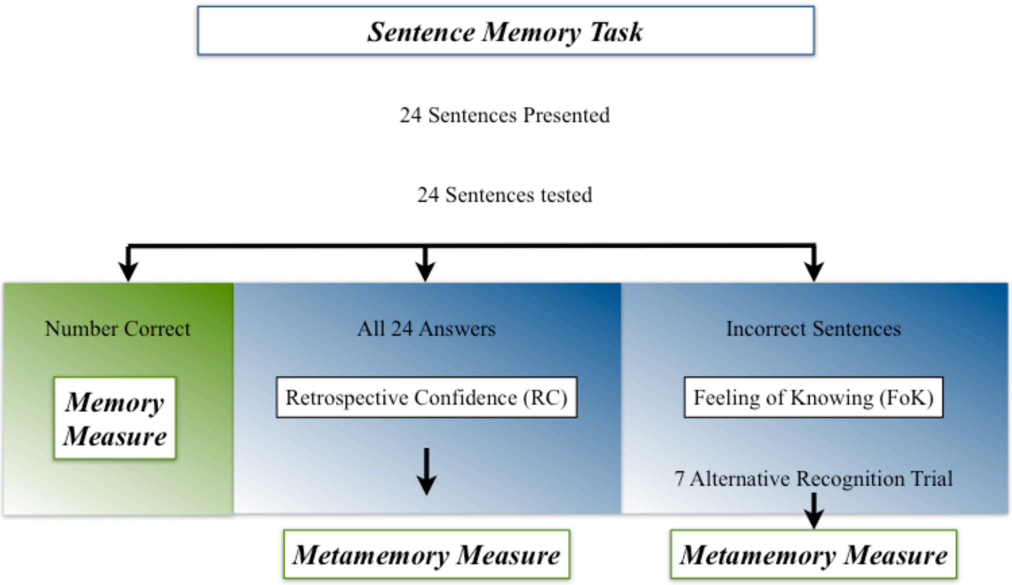


Figure 4.3 Summary of the derivation of both Retrospective Confidence Judgments and Feeling of Knowing judgment measures from the 24-item Sentence Memory test.

4.3.5.4. The Memory Functioning Questionnaire (MFQ; Gilewski, Zelinski & Schaie 1990, Zelinski, Gilewski & Anthony-Bergstone 1990).

Participants were asked to complete this measure by themselves, or, if they needed assistance in physically completing it, not to use the opinions of others to rate their memory. Large print versions and an audio version to accompany the questionnaire were offered to participants, where necessary; one of each was provided during the study. As this questionnaire had been provided in advance of meeting the researcher for the testing session of the study, all answers were checked for completion when the participant and researcher met. Incomplete questions were checked with participants, and answered only if they were incomplete as the result of error. Non-applicable items remained incomplete.

Three sets of scores were derived from the MFQ: the total score, the Forgetting While Reading score and the total score on the 10 item General Frequency of Forgetting (GFF-10) score. This GFF-10 score was the primary measure of memory self-efficacy. To retest the hypothesis of Randolph, Arnett & Freske (2004) who used Forgetting While Reading questions of the MFQ, these were also used as an index of memory self-efficacy.

4.4. Procedure

4.4.1. Ethical Approval

The study was given ethical sanction by the Brunel University School of Health Sciences and Social Care Research Ethics Committee. The main ethical issues addressed related to participant confidentiality, data protection, potential for distress and right to withdraw from the study. Confidentiality and data protection were addressed through an anonymisation process in recording data. All identifying information was retained as encrypted files for digitally held material, and in locked storage on the university site for other records. The potential for distress was highlighted in the participant information sheet, and right to withdraw from the study highlighted in both the information sheet and the consent form. A copy of formal ethical sanction is included in Appendix B.

4.4.2. Recruitment

Adverts were placed on the MS Society and MS Trust websites, local branch magazines and in MS Therapy Centres around Greater London. Potential participants were provided with a participant information sheet explaining the study, with follow-up contact arranged to answer any questions. The study information sheet outlined the nature of the research and the likely demands of taking part, including an indication of the level of reading/writing and speech abilities required by the task, likely length of testing session, as well as other inclusion criteria. A copy of the Participant Information Sheet is presented in Appendix C.

4.4.3. Obtaining Consent

Consent forms, one for signing and one for participants to keep, were provided with the Participant Information Sheet so they could be reviewed in advance of agreeing to participate. Once these had been provided, a follow up telephone call was made to answer any questions about the study, and if the participant remained interested they were enrolled, after clarification of the inclusion criteria.

Those enrolled to take part in the study were sent, and asked to complete, the Memory Function Questionnaire and to complete one consent form, the latter also to be signed by a witness. Appointments were made so that the testing session could take place within 2 weeks of provision of the Memory Function Questionnaire. This was necessary in order to maintain congruence between completing the questionnaire and completion of the Beck Depression Inventory, which asks participants to consider their mood over the previous 2 weeks (Beck, Steer & Brown, 1996). Participants for whom a within-two-weeks testing appointment could not be made were not sent the MFQ until within two weeks of the testing appointment.

Participants were contacted again in the days prior to the agreed testing to confirm availability; this also provided an opportunity to confirm that the Memory Function Questionnaire and consent form were completed in advance of testing. If testing was to be carried out away from the Brunel University site, a quiet room was advised, along with participants selecting their best time of day from the perspective of fatigue.

On the day of testing, the researcher took the completed consent form and Memory Function Questionnaire at the beginning of the testing session. At this time the Memory Function Questionnaire was checked for incompleteness and unanswered items were checked with the participant in order to reduce the occurrences of missing data. Those questions, for which the participant was unable to offer a response, remained incomplete. Where sought, presence of a companion during the testing session was permitted.

4.4.4. Structure of Data Collection

The testing session was structured in one of two ways to address potential facilitation or confounding effects of test order. Complete counterbalancing of subtests was not possible because of the necessity to have delayed recall on two tasks. The two testing orders are outlined in table 4.1. The aim was to preserve the same number of items between presentation and testing of the 24 sentences in the Sentence Memory Test and to allow for the same time period between presentation and testing on the Auditory Verbal Learning Test. Alternate participants received different test order administrations. Completing the

Beck Depression Inventory at the end of the testing session allowed for any distress to be addressed, as well as minimising the impact of that distress on performance during testing. The testing session was structured to last approximately 90-minutes and breaks were 20 minutes in length.

Table 4.1: Alternate versions of the testing session used in this study.

Test Order A	Test Order B
Sentence Memory Test – Presentation	Sentence Memory Test – Presentation
Digit Span Task	Brixton Spatial Anticipation Test
Symbol Digit Modalities Test	Hayling Sentence Completion test
Letter Number Sequencing Test	Auditory Verbal Learning Test Trials I-VI
Auditory Verbal Learning Test Trials I-VI	
Break	Break
Auditory Verbal Learning Test Trial VII	Auditory Verbal Learning Test Trial VII
Hayling Sentence Completion test	Digit Span Task
Brixton Spatial Anticipation Test	Symbol Digit Modalities Test
	Letter Number Sequencing Test
Sentence Memory Test – Recall & RCJ	Sentence Memory Test – Recall & RCJ
Sentence Memory Test –FoK	Sentence Memory Test – FoK
Beck Depression Inventory	Beck Depression Inventory

4.5. Analytic Strategy

The main issues with data analysis and management have been outlined in Chapter 3. The analytic strategy for this study had three stages. The first set of results sought to situate performance of the sample in respect of norms across the range of measures use. Analysis of performance by Memory and Executive Function subgroups and mood disorder subgroups was also carried out to address the findings and recommendations from previous work in the area (Beatty & Monson, 1991, Pannu & Kaszniak, 2005). This analysis used SPSS version 16.0 (SPSS Inc., 2007).

The second set of results presents the findings from submitting test results (observed variables) to confirmatory factor analysis in order to generate latent variables for the final set of analyses, the structural models; this included results of factoral assessment from previous studies. Confirmatory Factor Analyses (CFA) were carried out using AMOS 13 (Arbuckle, 2007) for Windows XP virtualised in Parallels version 3.0 for Macintosh.

Once factoral structure of these assessments was complete, regression-derived factor scores were created for the Beck Depression Inventory factors and for the Forgetting While Reading and GFF-10 scales. These factors scores were then used to represent the observed components of their respective latent variables. Regression-derived factor scores were generated using SPSS version 16.0 (SPSS Inc., 2007; Field, 2005)

Structural equation models (SEM) were specified based on previous work with neurologically impaired samples, and the theory base and findings regarding metamemory in non-neurologically impaired populations. These models were set out in Chapter 1: Literature Review. These analyses were carried out using AMOS 13 (Arbuckle, 2007) for Windows XP virtualised in Parallels version 3.0 for Macintosh. The analytic approach is outlined in figure 4.4

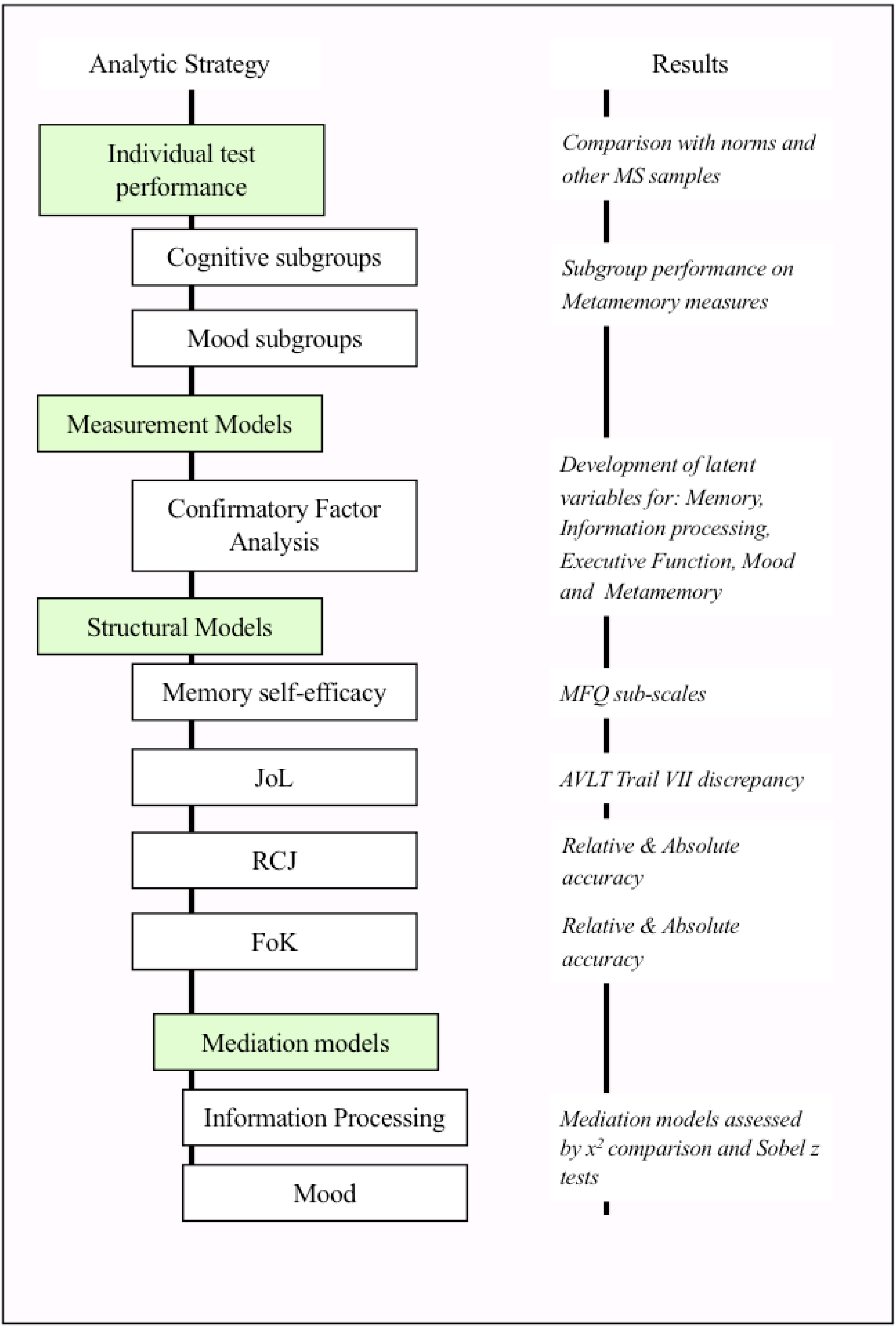


Figure 4.4. Summary of data analysis for the study.

Chapter 5: Sample Performance

5.1 Introduction

This chapter presents a description of the sample and their scoring on the range of cognitive, mood and metamemory measures used for the study. Analysis in this chapter aims to indicate performance on the range of measures used and present some indications of the relationships between the cognitive and mood variables and the metamemory measures used in the study. Full analysis of the relationships will be carried out in subsequent chapters presenting the measurement and structural models.

There are four aims in this chapter. First, to report the sample features, and compare with similar samples in the published literature in terms of demographics; second, to compare cognitive and mood measures to normed data; third, to report performance on the range of metamemory measures, and where possible to compare these to published literature. A final aim is to address the issue of heterogeneity in performance, by comparing subgroups of cognitive and mood function in the domains of MS subtype and metamemory measures. This will assess the findings of Beatty & Monson (1991) that subgroups of cognitive performance may be one way to understand differences in performance on metamemory measures. A summary at the end will discuss initial understanding of the results to aid subsequent structural models' interpretation.

5.2 Demographics, MS Subtypes and medications

Recruitment yielded 110 volunteers. Three people excluded themselves after receiving the participant information sheet. Additionally, one participant was excluded because of severe visual impairment, three were excluded because of past medical history (Chronic Fatigue Syndrome, history of psychotic illness, and anxiety disorder prior to MS diagnosis), and one person was excluded because they were outside reasonable travelling distance. Finally, after initial interest two people did not respond to follow up contact. After losses, and application of the inclusion criteria, the final sample consisted of 100 participants, all of whom were living in the community and had a confirmed diagnosis of MS. Figure 5.1 outlines the process of recruitment, retention and study losses. Of the 100 people, two were seen at Brunel University, and the remainder at home or at the MS Therapy centre they attended.

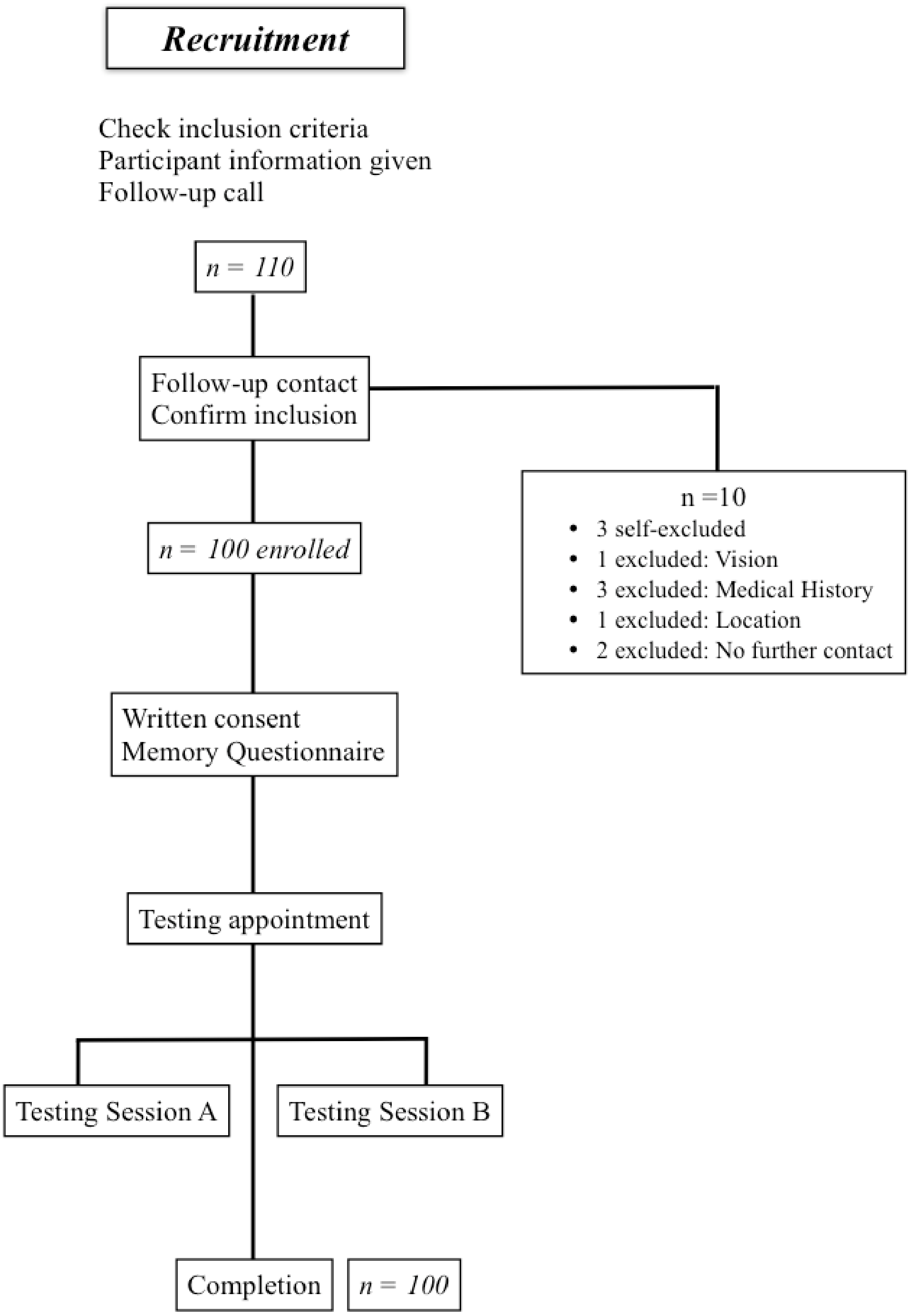


Figure 5.1. Summary of recruitment process, participant exclusions and losses. The final sample was 100 all of whom completed all components of the study.

The Sample

The mean age of the sample was 51.55years (SD = 11.61; range 25-84years; median age was 51 years) and the gender split was 22 (22%) male, 78 (78%) female, giving a ratio of 1:3.6. The mean age of male participants was 49.64 years (SD = 11.10) and mean age for female participants was 52.09 years (SD = 11.76).

Education

Mean number of years spent at school was 12.15 (SD = 1.41; Range 8-15 years). 50% attended full-time further education ranging from 1 year (n=3) to 7 years (n=1). Of those who spent some time in further education, 3 and 4-year periods were the most common (76%), and included education to degree level.

Employment

59 (59%) of participants were medically retired from paid employment (59% of females and 59.1% males); 16% were in full time employment (27.3% of men, 12.8% of women); 12% working part time (9% of men, 12.8% women). 6% were retired (all women); see Table 5.0 for status of remaining participants.

Table 5.0 Employment status of sample (n=100).

Employment Status	*%*
Full time	16
Part-time	12
Medically retired	59
Retired	6
Working < 1 day week	2
Volunteering	1
Full-time student	1
Self-employed	1
Not currently working	1
On sick leave	1
Total	*100*

MS Diagnosis.

For the sample as a whole, mean time since first symptoms was 21 years (SD = 10.66 years, range 42 years) and mean time since diagnosis was 13 years (SD = 9.37 years, range 39 years).

Participants were asked to give the type of MS given by their diagnosing doctor, and the subtype best characterised from their own experience. Table 5.1 summarises both subtype diagnosed, and experienced by participants. Slow progressive and progressive-relapsing types have been included as primary progressive forms of the illness, in line with the recommendations of Kremenchutzky et al., (1999). Hereafter, this five-category format of medically diagnosed subtype, including ‘none given’ will be used for comparisons.

Table 5.1: Multiple sclerosis Subtypes in the sample (n=100) Primary Progressive includes ‘slow progressive’ and ‘progressive relapsing’, which some consider types of primary progressive (Kremenchutzky et al., 1999).

MS Subtype	*Diagnosed*	*Experienced*
Relapsing/Remitting	32%	33%
Secondary Progressive	30%	35%
Primary Progressive	19%	24%
Benign	5%	8%
None Given	14%	
Total	*n =100*	*n=100*

Age & MS subtype

The mean age for receipt of a formal diagnosis of MS was 38.50 years (SD = 9.95), ranging from 18 to 65years and the mean reported age at first symptoms was 30.54 years (SD = 10.33), ranging from 5 to 61 years. This was a retrospective judgment about symptom onset.

A one-way ANOVA was used to test for differences between mean participant age for each of the MS subtypes (5 levels). Mean age for each of the MS subtypes differed significantly among groups, $F(4, 95) = 10.755$, $p<.001$. Tukey post-hoc comparisons of the MS subtype groups indicated that the Relapsing-Remitting group (M= 42.72 years; SD = 8.90) was younger than all other groups. Comparisons between all other groups revealed no statistically meaningful differences in respect of age.

Figure 5.2 shows the distribution of MS subtype for each of six decades of life covered by the sample. Visual inspection indicates a higher proportion of younger participants having a diagnosis of Relapsing Remitting MS (RRMS), with an increasing proportion of Secondary Progressive MS (SPMS) in groups of increasing age. There were no participants with benign MS over 66 years old. An increasing proportion of participants were categorised as primary progressive in older age

Time since diagnosis

Mean number of years since diagnosis was 13.10 years (SD = 9.36) ranging from a minimum of 1 year, to a maximum of 40. Mean number of years since first symptoms was 21 years (SD = 10.66), ranging from 3 to 45 years.

The time gap from first symptoms to diagnosis ranged from 0 to 34 years. The relative uniformity of mean time between first symptom and diagnosis, for each MS subtype, is demonstrated in the figure 5.3. A one-way ANOVA for MS subtype (5 levels) and time from first symptoms to diagnosis confirmed no meaningful differences between groups $F(4, 95) = 2.034$, $p = 0.096$ in respect of time from first symptom to diagnosis. This finding may relate to the 'wait and see' approach from initial symptoms, during which time dispersion of symptoms in time and space is required for a diagnosis.

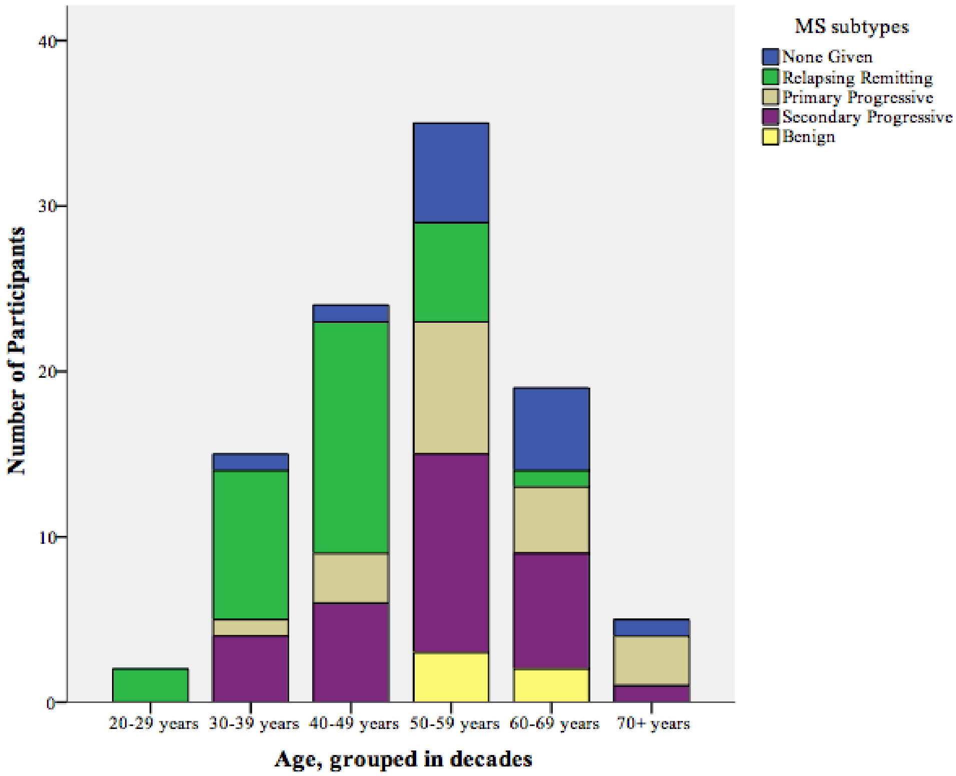


Figure 5.2 MS subtype summarised by age groups within the sample. The figure shows the proportions of MS subtype in each of 6 age groupings, with larger proportions of younger participants having a Relapsing-Remitting diagnosis and an increasing proportion of Secondary-progressive and Primary Progressive course in older age groups.

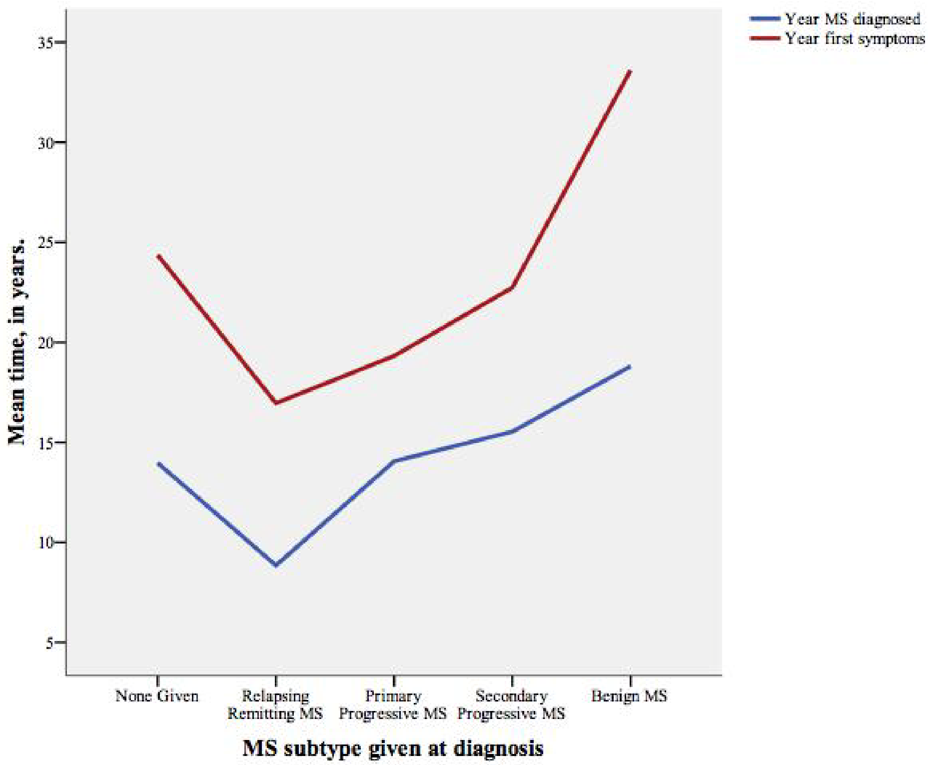


Figure 5.3. Mean time, in years since 1st reported symptom and formal diagnosis of MS. The gap between the two events (1st symptom, formal diagnosis) shows some uniformity, which may relate to the wait and see approach to diagnosis required to demonstrate dispersion of demyelination in time or place.

Medication use

The list of medications is organised according to the British National Formulary (BNF: Joint Formulary Committee, 2009). The treatment of neuropathic pain, which may involve the use of tricyclic antidepressants and antiepileptic medications, is also considered. Table 5.3 summarises medications in use by the sample; off-label use (e.g. of LDN), where reported, contributes to structuring the data.

Table 5.3: Medication use among the sample. Percent of sample currently using each is tabulated. Medications are categorised by their primary reasons for prescription according to the British National Formulary (Royal Pharmaceutical Society of Great Britain, 2009). Off-label use is also considered.

Medication category / Target Symptom	Percent of sample
Disease Modifying	
• Interferons	15
• Glatiramer Acetate	3
• Low-dose Naltrexone (LDN)	5
Relapse management	
• Prednisone	1
• Methotrexate	1
• Immunoglobin	1
Pain	53
Fatigue	3
Spasticity/neuromuscular	35
Continence	22
Depression & Anxiety	36
Sleep	4

The mean number of medications taken was 2 (SD = 1.61), ranging from none to six. 23% of the sample took none, 23% one, 22% two, 15% three, 9% four, 5% took five and 6% took six.

5.3. Comparison with other United Kingdom MS samples.

A sample of 136 community dwelling people with MS reported by O’Hara, De Souza & Ide (2000), also in the UK, and an economic survey (Kobelt, et al., 2006) with samples drawn in the UK using an MS charity, presents scope to compare demographic features of the sample in this study, with many similarities evident. Notably, Kobelt et al., (2006) suggests the mean annual cost per person, from a societal perspective is estimated at £30,263. Table 5.4 summarises the data from the two UK samples.

Table 5.4: Comparisons of three community samples of people with MS from the United Kingdom by age, gender, employment status and MS-related features. Employed include Full, part-time and self-employed employed. RRMS = Relapsing/remitting MS; CPMS = Chronic Progressive MS; PMS = Progressive MS; PPMS = Primary Progressive MS and SPMS = Secondary progressive MS.

	O'Hara, De Souza & Ide (2000) n=136	Kobelt et al., (2006) n= 2048	Current MS sample. n=100
Mean age in years (SD)	50 (10.75)	51.3 (10.7)	51.55years (11.61)
Gender	Male: 32%.	Male: 24.7%.	Male 22%.
Employment status	Employed 27% Volunteering 17% Not Working 56%	Employed 28% Retired due to MS 44% Not working 28%	Employed 29% Volunteering 1% Retired due to MS 59% Retired 6%
Years since diagnosis	Mean 12 (SD 8.6)		13 (SD 9.4)
MS Type	RRMS 47% CPMS 53%	RRMS 35% PMS 65%	RRMS 32% SPMS 30% PPMS 19% Benign 5% None Given 14%

5.3 Performance on Individual Tests.

Test Version.

Test version A was administered to 49% (n=49), and version B to 51% (n=51), of participants. No significant differences were found with an independent t-test between A and B Version groups for age $t(98) = -1.668$, p = 0.098, schooling $t(98) = -0.90$, p = 0.371, or higher education $t(98) = -0.88$, p = 0.380. A Pearson Chi-sq test of difference between test version (A or B) and MS subtype (5 levels) showed no significant difference between the test version given to each group $x^2(4, N = 100) = 4.45$, $p = 0.349$.

5.4.1. Memory

Auditory Verbal Learning Test (AVLT).

Comparisons against control performance are presented in table 5.5. These are based on norms published by Vakil & Blachstein (1993). For the AVLT, a number of scores are commonly used as measures of learning and memory (Lezak, Howieson & Loring, 2004).

The initial recall trial score (AVLT Trial 1), often used as an index of the impact of information processing capacities on initial acquisition, was lower for the MS sample $t(244) = 4.73$, $p < 0.001$ and performance on this item correlated with other proposed measures of information processing capacities in the sample; Symbol Digit Modalities Test, $r = 0.417$, $p < 0.01$, Digit Span Task $r = 0.292$, $p < 0.01$ and the Letter Number Sequencing task $r = 0.326$, $p < 0.01$. Total acquisition over all five learning trials (summed score), an index of learning rate, was also lower in the MS sample $t(244) = 5.61$, $p < 0.001$. 20-minute delayed recall, Trial 7, was lower for the MS sample $t(244) = 7.51$, $p< 0.001$.

Table 5.5. MS sample compared to a neurologically normal sample on the Auditory Verbal learning Test. Means are reported and standard deviation in brackets. Maximum score is 15 on all trials.

	Vakil & Blachstein (1993) n=146	*MS sample n=100*
Trial 1	7.73 (2.11)	6.47 (1.97)
Trial 5	13.30 (1.85)	12.05 (2.41)
Total (sum) trials 1 to 5	56.90 (9.03)	49.67 (11.12)
Trial7; 20 minute recall	12.23 (2.61)	9.45 (3.71)

20-Minute delayed recall cut-off score

Using the scores provided by Vakil & Blachstein (1993), the lower bound -1.5SD mark is 8.32. Therefore, the cut-off scores for normal performance on delayed recall are scores > 8, classifying 36% (n=36) as impaired on this task.

A number of studies suggest memory is impaired in MS and performance on this test accords with findings of Bravin et al,. (2000), who had a younger, but community dwelling, sample recruited form MS Society membership in Australia (Mean 45.75 years, SD = 9.68). In that study, Trial 1 performance was 5.38 (2.08) and Trial 7, 6.77 (4.16), suggesting the sample in this study was less impaired.

Sentence Memory Test

There are no norms available for this test, as it was designed for the study. Performance is reported based on 60-minute delayed recall on 24 sentences completions. For the 100 participants, the number of sentences completed correctly ranged from 0 to 22, with a mean = 9.50 (SD = 4.08).

Recognition performance, on the unrecalled items, with a forced choice 7-alternative format was mean 47.33% (SD 20.61), ranging from 5% to 100%. No differences in recognition performance were detected in the MS subtypes.

5.4.2. *Executive Function*

Hayling Sentence Completion Test

The three individual measures of the Hayling Sentence Completion Test - Response time Sections A and B, and Errors Section B (Burgess & Shallice, 1997) - can be individually compared to scaled scores, with a score of 6 being equivalent to the 50th percentile for each. A combined scaled score can also be computed taking account of trade–offs made between reaction time and accuracy. This overall score is again scaled so that 6 is the average score. Table 5.6 indicates sample performance on Response Time Section A, Response Time Section B, Errors Section B and Overall Score, against the normed mean of 6, which represents an ‘average score’ or, performance at the 50th percentile level (Burgess & Shallice, 1997).

Table 5.6: Mean Scaled Scores (SD) on Hayling Sentence Completion test of MS sample and a control sample reported by the test authors (Burgess & Shallice, 1997). n.s. indicated no reliable difference between norm and MS sample; for all other comparisons the MS sample were lower in response time, had fewer errors and a lower overall scaled score.

	Response Time Section A (n.s.)	Response Time Section B	Errors Section B	Overall Scaled Score
MS sample n=100	5.14 (1.2)	4.76 (1.8)	5.65 (2.2)	4.87 (2.0)
Controls n=71	5.8 (1.0)	5.9 (1.0)	6.4 (1.7)	6.1 (1.6)

A t-test of MS sample means, indicated they differed from controls on *Response Time Section B t*,(169) = 4.83, p < 0.001, *Errors Section B t*,(169) = 2.41, p <0.05 and on the *Overall Scaled Scores t*,(1969) = 4.30, p < 0.001. Response times for Section A (easy sentence completion) was not different between the MS and control samples, perhaps suggesting that this was not sensitive to information processing deficits.

According to the authors (Burgess & Shallice, 1997), scaled scores of 4 or below are categorised as Low Average (4 = 10th percentile), Poor (3 = 5th percentile), Abnormal (2 = 1st percentile) or Impaired (1 = <1st percentile). The number of participants scoring at these levels is presented in Table 5.7. All those scoring at or below the 5th percentile (scores of 3 or less) are considered here to be impaired, as it tallies closest with the -1.5SD cut off used for other measures.

14% of the sample were impaired on Response Time Section A (response initiation, simple condition), 20% were impaired in Response Time, section B (complex condition) and 21% impaired in the number of errors generated in Section B. Overall scores on the sentence completion task suggested 23% can be considered impaired. Since the overall score is a composite of the two response times and error scores, some participants, who were impaired on more than one aspects of the test can be included more than once. Table 5.7 summarises performance.

Table 5.7: Summary of scoring on the Hayling Sentence Completion Task: Numbers of participants scoring ≤10th percentile (a score of 4 or less) on each component of the Hayling Sentence Completion Task. Note: Overall Scaled score is a conjunction of the other 3 scores, and the same participant may be represented more than once if their score on two categories was ≤10th percentile. Shaded cells are those participants deemed impaired because their overall performance is below the -1.5 SD cut-off.

	Response Time Section A	Response Time Section B	Errors Section B	**Overall Scaled Score**
Low Average (10th percentile)	5	18	2	8
Poor (5th percentile)	12	6	3	**6**
Abnormal (1st percentile)	1	1	12	**2**
Impaired (<1st percentile)	1	13	6	**15**

Brixton Spatial Anticipation Test

Mean number of errors on the Brixton Spatial Anticipation Tests was 18.13 (ranging from 4 to 32; SD = 6.32) with n = 98. Two participants' data was missing because of examiner error. Scaled scores for n = 98, and associated classification, are outlined in Table 5.8. 23% were classified as impaired.

Table 5.8: Numbers of participants scoring ≤ 10th percentile (a score of 4 or less) on the Brixton Spatial Anticipation Test. Impaired scores are contained in shaded cells and are equivalent to < -1.5SD; n=98.

	Overall Scaled Score (n=)
Low Average (10th percentile)	7
Poor (5th percentile)	6
Abnormal (1st percentile)	14
Impaired (<1st percentile)	3

In comparison to the standardisation sample (mean 16.0 errors, SD = 5.7), the MS sample had a higher number of errors t,(169) = 2.27, $p < 0.05$. Setting the cut-off for impaired scores at the -1.5 SD, 23% (n = 23) of the sample is considered impaired.

5.4.3. Information Processing.

Letter-Number Sequencing Test (LNS)

Mean score on this test was 10.43, (SD = 2.57) with a maximum score available of 21. The range of scores was 17 (3-19). For the purposes of classification, age-scaled score equivalents were computed for each participant using the UK standardization sample for the Wechsler Adult Intelligence Scale, (WAIS-III; Wechsler, 1997b). For the sample in this study, the age scaled mean score was 10.67 (SD = 2.78). 21% (n = 21) of participants fall between the 1st and 2nd SD below the mean of age scaled scores, with 2% (n=2) below the -2 SD point. 23% (n=23) were therefore classified as impaired on this task.

Digit Span Task - Forwards

Mean score on the Digit Span Task was 10.0 (SD = 2.14). The range of scores was 11 (6-16). Norms, for this task, are available from Strauss, Sherman & Spreen (2006:240). A standardisation sample of 540 non-neurogically impaired adults was used, and scores are age stratified. Table 5.9 compares performance between the norm group and the MS sample in this study (n=100), and indicates the proportion of the MS sample classifiable as impaired using ≤ -1.5 SD as the cut off. The number classified as impaired, 3%, would be expected within the ‘normal’ population, given that ≤ -1.5 SD theoretically represents about 7 percent of the normal distribution. This result therefore suggests average performance of the sample on this task.

Table 5.9: Age stratified comparison of performance on the Digit Span Forwards Task. Norms from Strauss, Sherman & Spreen, (2006) are used. Scores are analysed against a cut-off of ≤ -1.50 SD from the mean of the standardisation sample.

Age Range	Norms M (SD)	MS Sample M (SD)	Cut-off Score at -1.50 SD	% MS sample impaired i.e. ≤ -1.50 SD
20-39	11.7 (2.5)	10.18 (1.74)	7.95	0
40-49	10.6 (2.2)	9.92 (2.10)	7.30	1
50-59	10.5(2.4)	10.31 (2.31)	6.90	0
60-69	10.2(2.1)	9.37 (2.36)	7.05	2
70-79	10.4(2.5)	9.75 (1.70)	6.65	0
80-89	9.2 (2.2)	11 (0) n=1	5.90	0

Symbol-Digit Modalities Test

Mean score (total correct, in 90 seconds) was 45.57 (SD = 12.40). Norms for the SDMT are stratified by six age bands and by education in the test manual (Smith, 1982). Table 5.10 summarises the performance of the sample against these stratified norms. There was no statistical difference, in an independent t-test, in scores between the ≤ 12 years schooling and the 13+ years schooling groups for the MS sample, t (98) = -0.69, p = 0.50.

Table 5.10: Mean (SD) scoring on the SDMT stratified by age and level of schooling, using norms from Smith (1982).

	12 years school or less		*13 or more years school*	
Age Range	*Norms*	*MS sample n=52*	*Norms*	*MS sample n=48*
25-34	60.57 (9.14)	52 (25.46)	65.71 (11.64)	56.67 (10.50)
35-44	59.87 (10.49)	51.82 (11.88)	60.95 (11.32)	51.86 (14.11)
45-54	53.91 (10.40)	47.57 (9.65)	58.31 (8.67)	42.81 (12.587)
55-64	49.03 (9.03)	39.00 (9.10)	54.47 (8.93)	45.33 (11.45)
65+	42.05 (11.26)	40.20 (12.83)	52.89 (13.54)	40.17 (8.57)

The test authors suggest that ‘*1 to 1.5SD below the mean should be considered suggestive of cerebral dysfunction*’ (Smith, 1982:14), meaning it should be possible to classify participants into ‘Low’ (approximately -1SD) ‘Moderately Low (approximately -1.5SD) and ‘Very Low’ (approximately -2SD) groups. Table 5.11 summarises the results of this analysis. At the ≤ -1.5SD level (moderately- and very- low scorers) 28% of the sample would be considered impaired; wider parameters, such as including ‘Low’ scorers, would capture 48% of the sample.

Table 5.11: Classification of performance by the MS sample (n = 100) on SDMT, using age and education-stratified norms of the standardisation sample (Smith, 1982).

Age group	*Low*	*Moderately Low*	*Very Low*
25-34	0	1	1
35-44	5	3	4
45-54	7	2	7
55-64	4	2	6
65+	4	1	1
Totals	*20%*	*9%*	*19%*

One limitation for these published norms (Smith, 1982) is that they are based on verbal scoring following a written attempt at the task. This may inflate the norms somewhat as prior experience may facilitate subsequent oral performance, at least in neurologically unimpaired samples (Barr, 2001).

Sheridan et al., (2006) investigated normative performance in community dwelling neurologically normal volunteers on the SDMT. They provide aggregate scores from a range of previous studies, including oral-only norms. These aggregated means, stratified by age group, were compared with the MS sample, yielding scores in 36 participants (36% of sample) that are considered indicative of impairment, using the lower bound 1.5SD cut off. Table 5.12 presents a summary of this data and comparisons for each of the age-stratified MS groups against the same age group normative data.

Table 5.12: Statistical comparison of age-stratified performance on the *oral-only* version of the Symbol Digit Modalities Test to derive indications of impaired performance among the MS sample. Sheridan et al (2006) presents oral-only norms for community-dwelling neurologically unimpaired samples. For each age group, the MS sample is reliably slower in their performance on an independent t-test

Age Group	Sheridan et al (2006) mean (SD)	MS sample mean (SD)	Comparison t test	-1.5SD cut-off value	Number (and %) of MS impaired
< 30 years n=6	69.40 (10.6)	58.17 (15.70)	p < 0.05	≤ 53	2 (30%)
30-55 years n=59	59.50 (9.2)	47.59 (11.60)	p < 0.001	≤ 45	26 (44%)
> 55 years n=35	47.30 (11)	40.00 (10.73)	p < 0.001	≤ 30	8 (23%)

5.4.4. Mood

Beck Depression Inventory 2nd Edition (BDI-II)

Median score on the BDI-II was 9.00, (IQR = 12; Range 0 - 43); using the cut-off given by Beck, Steer & Brown (1996), the most frequently endorsed symptom level was Minimal/None (mode = 0). There was no significant difference between male and female scores (t (98) = 0.906, p = 0.302). The levels of self-reported depression are reported in Table 5.13.

Table 5.13. BDI-II Ratings of sample (n=100).

Depression symptom level	*% of sample*
Minimal/none	64
Mild	18
Moderate	13
Severe	5
Total	*100*

A one-way ANOVA by MS subtype (5 levels) indicated no differences in BDI-II scores between the MS subtypes F(4, 95) = 2.126, p = 0.084. Figure 5.4 summarises self-reported depression by each MS subtype.

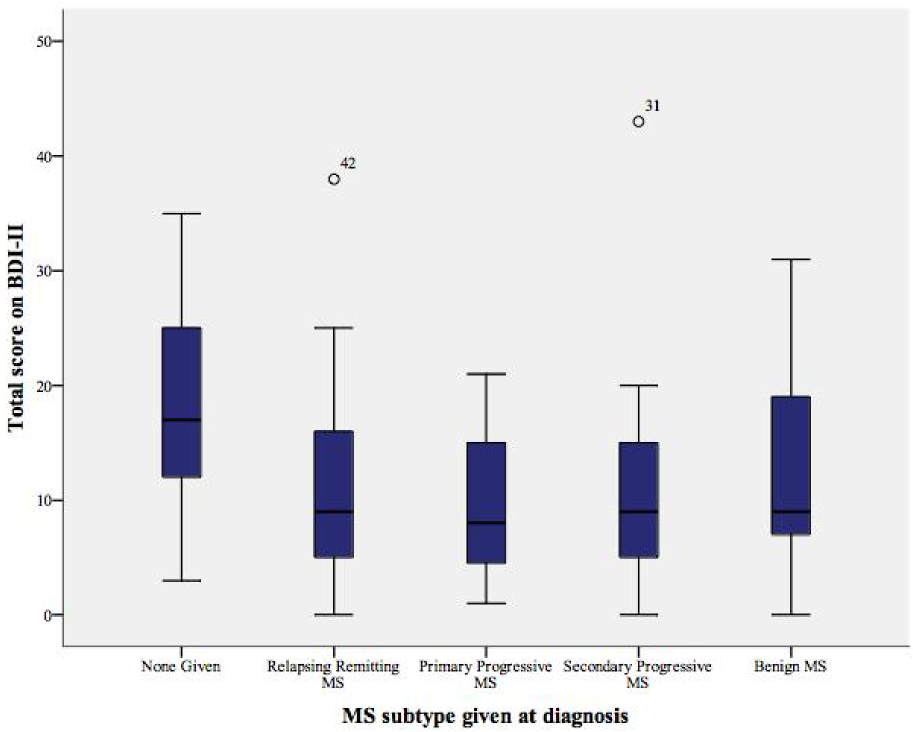


Figure 5.4. Total score on the Beck Depression Inventory 2nd edition grouped by MS subtype. Maximum score is 63 and lower scores indicate less self reported depression symptoms.

5.4.5 The Memory Function Questionnaire, total score

Higher scores in the MFQ indicate less frequent problems, less seriousness attributed to problems when they do occur, same or better memory compared to the past, and less use of mnemonics. Missing data (unanswered questions) was present for 11 participants in calculating the total MFQ score. Because of the distribution of unanswered questions, for the four subscales sample numbers varied from 92 to 100. For the total, mean score was 237.42 (SD = 54.89), n=89.

Reasons for not answering questions were often given on the questionnaire and typically related to personal relevance - questions about forgetting while public speaking, taking tests and chores were the most common unanswered questions, leading also to non-answering of the same items when asked about how serious forgetting was rated, when it happened. Other unanswered questions related to forgetting while reading items not being applicable; typical reasons included difficulties holding a book/newspaper and reading difficulty relating to visual problems. These missing items can be considered generally to have a pattern to their missingness in that they relate to MS disability.

Comparison, using t-tests, of age stratified scores for the MS group and the standardisation sample, indicate that people with MS reported a higher frequency of memory problems, $t (868) = 5.58$, $p < 0.001$, and considered them more serious when they did occur, $t (870) = 6.00$, $p < 0.001$. The MS sample reported better past memory performance, $t(876) = 9.89$, $p < 0.001$, that is, more decline in memory over time. Use of mnemonics was reported as higher among the MS sample, $t (876) = 6.89$, $p < 0.001$. Table 5.14 indicates the age-grouped performance of the normalisation sample and the MS sample, with the questionnaire divided into its four constituent subscales.

Table 5.14: Age stratified comparisons of the Memory Function Questionnaire standardisation sample Gilewshi, Zelinski & Schaie, (1990) across each of the four sub-scales: General Frequency of Forgetting, Seriousness of Forgetting, Retrospective Functioning and Mnemonic usage. Where there is incomplete data for a scale this is indicated by the number of MS participants that contribute data to the comparison. Lower scores indicate more memory complaint, more seriousness of failures, greater decline from past memory (retrospective functioning) and higher reported mnemonic usage. M = mean, SD = standard deviation

Age ranges.	General Frequency of forgetting (33 items)		Seriousness of forgetting (18 items)		Retrospective functioning (5 items)		Mnemonics usage (8 items)	
	Norm M(SD)	MS Sample M(SD)	Norm M(SD)	MS Sample M(SD)	Norm M(SD)	MS Sample M(SD)	Norm M(SD)	MS Sample M(SD)
16-29 years	162.26 (22.19)	143.50 (64.35) n=2	77.10 (20.13)	42.00 (10.00) n=2	22.96 (5.08)	17.00 (4.24) n=2	32.47 (9.59)	21.00 (4.24) n=2
30-49 years	166.51 (26.21)	130.88 (35.04) n=37	85.58 (21.59)	69.74 (23.48) n=38	20.90 (4.75)	11.33 (6.072) n=39	29.42 (8.43)	22.51 (8.75) n=39
50-59 years	155.99 (25.53)	141.24 (35.84) n=33	83.75 (20.02)	69.21 (22.26) n=33	17.70 (4.62)	15.05 (4.99) n=35	30.49 (8.89)	23.05 (10.12) n=35
60-69 years	152.13 (28.12)	139.13 (32.36) n=16	85.52 (20.09)	69.64 (20.01) n=17	18.31 (5.68)	13.21 (4.90) n=19	30.87 (9.44)	26.21 (12.26) n=19
70-79 years	148.72 (29.12)	147.00 (30.05) n=4	83.55 (20.89)	91.50 (17.89) n=4	18.18 (5.74)	12.00 (3.93) n=5	29.70 (10.26)	23.00 (7.28) n=5
Totals	*154.63 (27.83)*	*137.01 (34.73) n=92*	*83.48 (20.56)*	*69.87 (22.35) n=94*	*19.10 (5.69)*	*13.14 (5.51) n=100*	*30.40 (9.52)*	*23.40 (9.87) n=100*

Memory self-efficacy

The primary measure of memory self-efficacy in this study is the 10-item scale drawn from the 33-item General Frequency of Forgetting scale (GFF), itself drawn from the Memory Function Questionnaire (MFQ, Lane & Zelinski, 2003; Zelinski & Gilewski, 2004). The relationship between the different scales was set out in Chapter 4

The norms presented are for both 33-item and the reduced 10-item self-efficacy scale (Zelinski & Gilewski, 2004, using two samples). Missing data was present for ten participants in calculating the total MFQ score, 8 for the 33-item General Frequency of Forgetting scale and for 4 in calculating the 10-item self-efficacy scale. Descriptive statistics for the MFQ subscales are calculated on available cases, so samples range from n=92 to n=100.

Two samples reported by Zelinski & Gilewski (2004) with a mean age of 69.6 and 68.9 years were compared with the MS sample on the 10-item GFF scale. Independent t-tests, comparing samples 1 and 2 to the MS sample demonstrated reliable differences between the MS group and Sample 1, $t(663) = 6.54$, $p < 0.001$ and Sample 2, $t(581) = 5.92$, $p < 0.001$. The MS sample scored lower, indicating more reported problems with memory in day-to-day functioning, despite a lower age range. Given differences in age, with the norm samples being older by about 14 years, this comparison may underestimate the differences between the two groups. Performance of each of the three groups is summarised in Table 5.15

Table 5.15: Comparisons of two unimpaired samples and the MS sample on the 10-item General Frequency of Forgetting scale (Zelinski & Gilewski, 2004). The number of participants contributing data to each comparison is indicated where there is missing data. n = number of participants in each group; *SD* = standard deviation.

	10-Item General Frequency of Forgetting scale		
	n	Mean	*SD*
Sample 1	565	4.7	*0.9*
Sample 2	483	4.7	*1.0*
MS sample	96	4.05	*1.02*

Memory Functioning Questionnaire scoring in other MS Samples.

Randolph Arnett & Freske's (2004) sample of MS participants also completed the MFQ. Only the items relating to Forgetting While Reading (FWR) are reported for that study, with a mean FWR score of 51.4 (SD = 12.9); here the FWR mean was 44.62 (SD = 14.65). An independent sample t test suggests that the samples' ratings were different with the MS sample in this study (n=100) scoring lower on forgetting while reading items t, (146) = 2.74, $p = 0.007$. Lower scores on this item indicate more frequent reports of forgetting while reading.

Associations between MFQ and Memory performance.

Each of the four MFQ subscales, and the shortened GFF-10 were submitted to bivariate correlation analysis with the two memory indices for the study - 20-minute delayed recall on the Auditory Verbal Learning Test and the Sentence Memory Test, 60-minute delayed recall. The full table of correlations is presented in Table 5.16

The substantive questions were of relationships between self-reported memory and actual performance. Here, there was evidence of association between General Frequency of Forgetting scale and performance on the Sentence Memory Test (Pearson's $r = 0.24$, $p < 0.05$) suggesting those with better memory performance on the Sentence Memory Test endorsed less frequent memory problems in daily life (higher GFF scoring). This finding concurs with a small association found by the developers of the GFF-10 scale (Zelinski & Gilewski, 2004) that recall performance explained small amounts of variance, even with individual factors controlled.

20-minute delayed recall on the Auditory Verbal Learning Test (AVLT 7) did not correlate with self-reported memory performance (10-item GFF), but there was a significant negative correlation with reported frequency of Mnemonic use ($r = -0.211$, $p < 0.05$) suggesting that better performers on the AVLT reported more frequent use of mnemonics.

Table 5.16: Pearson correlations between six Memory Function Questionnaire (MFQ) scales and memory measures: AVLT 7 and Sentence Memory test (SMT) total recall. ** = Correlation is significant at the 0.01 level (2-tailed); * = correlation is significant at the 0.05 level (2-tailed). GFF = General Frequency of Forgetting; SF = Seriousness of Forgetting; RF = Retrospective Functioning; MU = Mnemonics Usage; FWR = Forgetting while Reading

	10-item GFF	SF MFQ	RF MFQ	MU MFQ	FWR MFQ	AVLT 7	SMT Recall
10-item GFF	1.000						
SF MFQ	.424**	1.000					
RF MFQ	.398**	.326**	1.000				
MU MFQ	.220**	.242**	.081	1.000			
FWR MFQ	.790**	.371**	.383**	.253**	1.000		
AVLT 7	.051	.099	-.024	-.211*	-.051	1.000	
SMT Recall	.243*	.000	.009	-.038	.114	.555**	1.000

5.4.6. Task-based Metamemory Judgments.

Judgment of learning

This was measured as predicted minus actual performance (i.e. discrepancy) for the delayed recall trial of the Auditory Verbal learning Test (Trial 7). Figure 5.5 presents the performance of the sample over all trials. Trial 1 indicates judgment prior to initial acquisition, having only being told there were 15 words to remember. Both Trials 3 and 5 have been proposed to be the maximum acquisition point (Spreen & Strauss, 1998; Lezak, Howieson & Loring, 2004) and Trial 7, delayed recall. In advance of each trial, participants were asked to estimate how many of the same list of 15 words they predicted they would remember. Trial 6 was a new list.

Trial 7 is the key measure; there was a mean discrepancy of -2.7items (SD = 2.63) in the prediction. It suggests that actual recall was better than predictions made for the 20-minute delayed trial. Mean (SD) discrepancy for Trial 1 was +0.35 (2.66), Trial 3 was -2.21 (2.41) and trial 5 was -2.02 (2.42). Comparisons suggested Trial 1 was reliably different from the others - on paired t-tests all p-values were < 0.001. Trial 7 was also statistically different from Trial 5 $t(99) = 2.148$, $p = 0.031$. These results suggest some underconfidence across repeated trials 2 to 5 of the word list. Of interest here is the suggestion that the delayed Judgment of Learning (Trial 7) was more inaccurate (larger discrepancy) than the earlier trials. This is an unusual finding, given the weight of evidence suggesting that delaying a Judgment of Learning tends to increase accuracy (Dunlosky & Metcalfe, 2009).

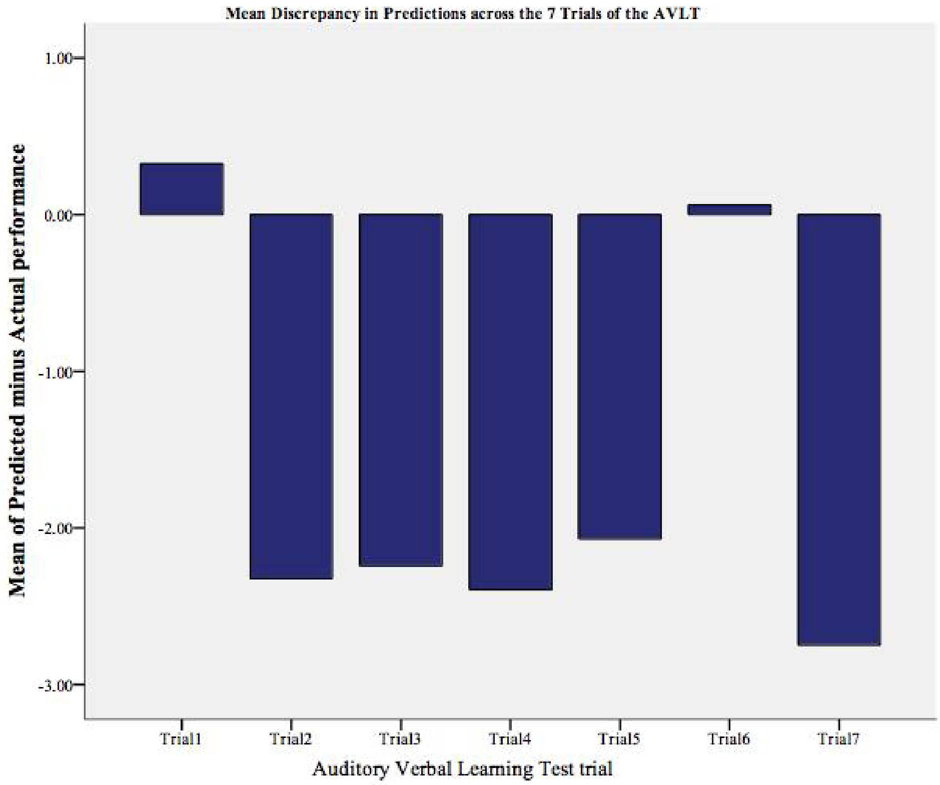


Figure 5.5. Mean discrepancy between predicted and actual recall performance on each trial of the Auditory Verbal Learning Test. Discrepancy was calculated by subtracting actual performance form predicted performance on a 15 item word list in advance of presentation and recall attempt for each trial. Positive values indicate expectations that are over confidence, negative scores underconfidence. Trail 7 was tested after a 20-minute delay without presentation of the list to the participant.

Retrospective Confidence Judgment.

RCJ Relative Accuracy

This judgment was made using a choice of four ratings after retrieval of an answer for each of the 24 Sentence Memory Test items. Rating use tended to be full, with 4% using only two ratings of confidence, 23% using three and 73% using all four. There were two missing datasets for this measure, so that n=98. Don't know responses for retrieval confidence were considered a rating of 1 (Low confidence / guess). 49% of respondents had none, 10% had one, with the remaining 41% having up to 16.

Because of near full use of the rating procedure, the gamma correlation was calculated using the method provided by Schraw (1995). The mean Retrospective Confidence gamma value was 0.89 (SD = 0.197; maximum = 1.0, minimum = -0.302). 42% of participants scored the maximum 1.0, indicating perfect resolution on this measure, that is, confidence in recalled answers matched success in recall. Remaining scores ranged from 0.98 to -0.302. A One-way ANOVA of mean gamma correlation by MS subtype (5 levels) showed no significant group differences, $F(5, 94) = 0.543$, $p = 0.744$. The distribution of gamma scores is presented in figure 5.6 and suggests a highly skewed distribution, but also high levels of accuracy, indicated by the frequency of values at or near 1.0.

RCJ Absolute Accuracy

The proportion of High Confidence judgments with correct retrieval, and of Low Confidence with correct retrieval were also taken as measures of absolute accuracy. The first is proposed as an indicator of accuracy, the second an indicator of inaccuracy.

For the sample, the mean % High confidence correct was 76.23% (SD = 21.22), ranging from 0 to 100%. For the % Low confidence correct, the mean was 12.01% (SD 12.56) with a range from 0 to 74%. In terms of RCJ calibration there were high levels of accuracy (76%) and low levels of inaccuracy (12%) for the sample as a whole on this task.

MS subtype did not discriminate performance on a one-way ANOVA for either %High Confidence correctly recalled $F(4,95) = 0.145$, $p = 0.965$ or for % Low Confidence correctly recalled $F(4,95) = 0.339$, $p = 0.852$.

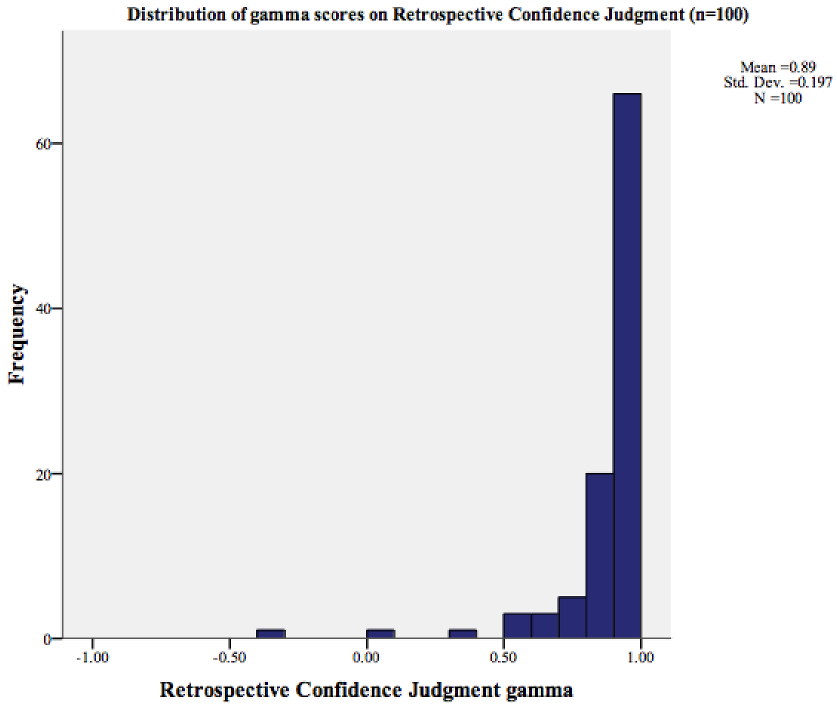


Figure 5.6. Distribution of Retrospective Confidence Judgment gamma scores, which can range from +1.0 to -1.0. Higher scores indicate more accuracy. The graph demonstrates the high levels of accuracy in this judgment, denoted by scores close to and reaching a value of 1.0.

Feeling of Knowing Judgment

FoK Relative Accuracy

Relative accuracy, or resolution was calculated using the non-parametric gamma correlation based on the ranking process outlined in Chapter 3: *Development of Statistical Methods.* This was carried out because of incomplete use of the 4 ratings offered.

The mean Feeling of Knowing gamma value was 0.18 (SD = 0.38), which ranged from -1.00 to +1.00 as the correlation allows, with a median of 0.195 and mode of 0.50. 29% of participants had values less than zero. Values towards + 1 indicate higher levels of accuracy and values towards -1 suggest systematic inaccuracy in this task. There were no group differences relating to MS subtype on a one-way ANOVA, F (5, 94) = 0.718, p = 0.611. FoK gamma scores are summarised in Figure 5.7.

As a comparison, the findings of Beatty & Monson (1991) suggest the mean sample gamma for controls was 0.24, (SD 0.32; n=22; mean age: 49 years); here it was 0.18; (SD 0.38; n=100), which is statistically similar (t (120) = 0.6882, p = 0.49).

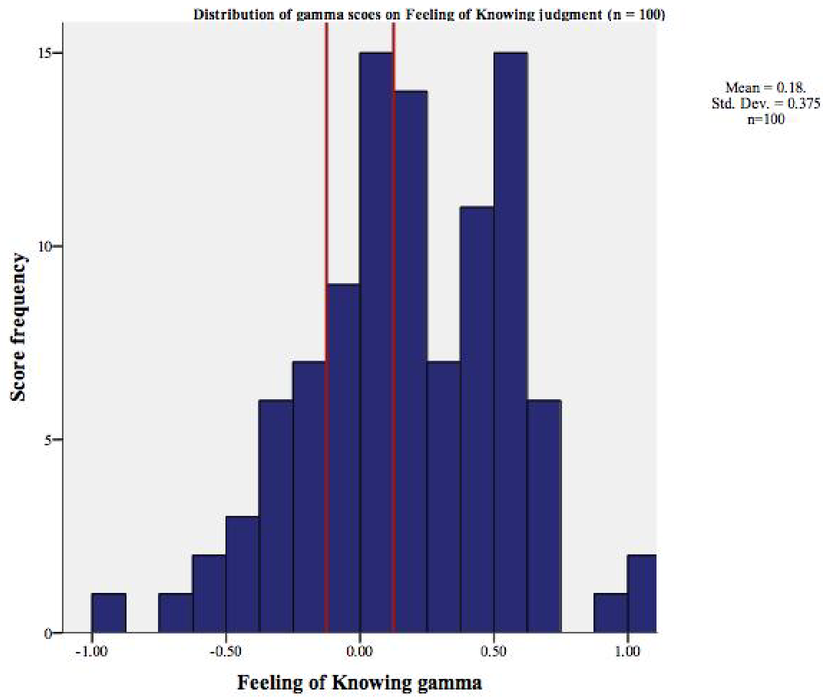


Figure 5.7. Distribution of Feeling of Knowing gamma scores. Values range from +1.0 to -1.0. Higher scores indicate more accuracy. The graph demonstrates low levels of accuracy, with only 2 scores reaching 1.0. Scores between the red lines (absolute values 0.125) indicate chance performance on the 7-alternative forced choice recognition test.

FoK Absolute accuracy

Absolute accuracy, or calibration was calculated by examining the % of Sentence Memory Test items correctly recognised for High FoK and Low FoKs separately. Variable use of the four available ratings led to some difficulties with the calculation of calibration of accuracy with the FoK, an issue that has been reported in a previous study with MS samples (Beatty & Monson, 1991). It has been noted more generally, that there may be restricted use of the full ratings, possibly indicating sensitivity to task difficulty, leading to what has been termed mid-point anchoring (Moulin, 2002). A count of rating use indicates that 10% (n=10) used only one rating 8 of whom used only a low ranking, 31% (n=31) used two (mostly rating 2 and 3), 42% (n=42) used three ratings and 17% (n=17) used all four ratings. This is in contrast to the same rating used in the RCJ's, within the same task, where 73% used all 4 ranks.

Aside from the potential utility of examining rating use, it also made it difficult to decide whether the use of two low ratings (e.g. 2 and 1) were to be interpreted as Low FoK or whether one could be considered High and the other Low. Initially, ratings were collapsed into High (4 & 3 ratings) or Low (2 & 1 ratings). In calculating the proportions correct, this generated missing data because division by zero was attempted, caused by rating non-use. Excluding the missing data provided the following results; %High FoK correct (n=97); Mean = 52.84 (SD 24.95). For % Low FoK correct (n=80); mean = 38.92 (SD 26.66).

In attempting to recover data, using a method outlined by Beatty & Monson (1991), ratings that were in the same end of the scale, e.g. only 4 and 3, or only 1 and 2 were split to reflect High or Low. For the derived %High FoK, subsequently recognised correctly; the mean score was now 54.43% (SD 26.11) n=98 and for the %Low FoK judgments subsequently correct; the mean was now 38.51% (SD = 27.435) n=92. This reduced data loss and maintained global characteristics of the data.

Comparison of the means by t-test showed no reliable difference between the two %High FoK $t(193) = 0.4347$; $p = 0.66$, or the % Low FoK correct datasets, $t\ (170) = 0.099$; $p = 0.92$ so the process was considered acceptable. No statistical differences were found between mean of % High FoK correct, or % Low FoK correct, for any of the MS Subtypes.

5.5 Investigating Heterogeneity of performance in MS

Pannu & Kaszniak (2005) and Beatty & Monson (1991) suggest that consideration be given to subgroups of performance among potentially heterogeneous samples, such as participants with MS or traumatic brain injury. Heterogeneity could relate to the subtype of MS, but comparisons carried out suggest not, in this sample.

Instead, Beatty & Monson (1991) attempted to address the potential confound of heterogeneous performance among people with MS by grouping performance according to cognitive performance, specifically memory and executive function. Here, consideration is given to Memory (AVLT7 score), and Executive Function (Total scaled score for Hayling test and/or Brixton spatial anticipation). Sentence Memory Test performance has no normative data from which to derive impaired memory performance so was not included. An additional factor potentially increasing heterogeneity is mood disorder.

For each, a simple division was placed between those classified as impaired based on the foregoing analyses - scores of above 8 on the AVLT 20 minute delay, scaled Hayling and scaled Brixton scores at or below the 5th percentile according to the test authors' recommendations (Burgess & Shallice, 1997). Executive function was therefore classified as impaired in either Hayling or Brixton, or both, if scores were at or below the 5th percentile. Four sub-groupings were generated; *No Impairment* on either memory or executive function, *Memory only* impairment, *Executive Function only* impairment, and finally, *Memory and Executive* impairment. Figure 5.8 summarises the proportion of MS subtype representing each of the cognitive subgroups. Of note are the high proportions of both primary progressive and secondary progressive MS classified as cognitive impaired, compared to the relapsing-remitting course.

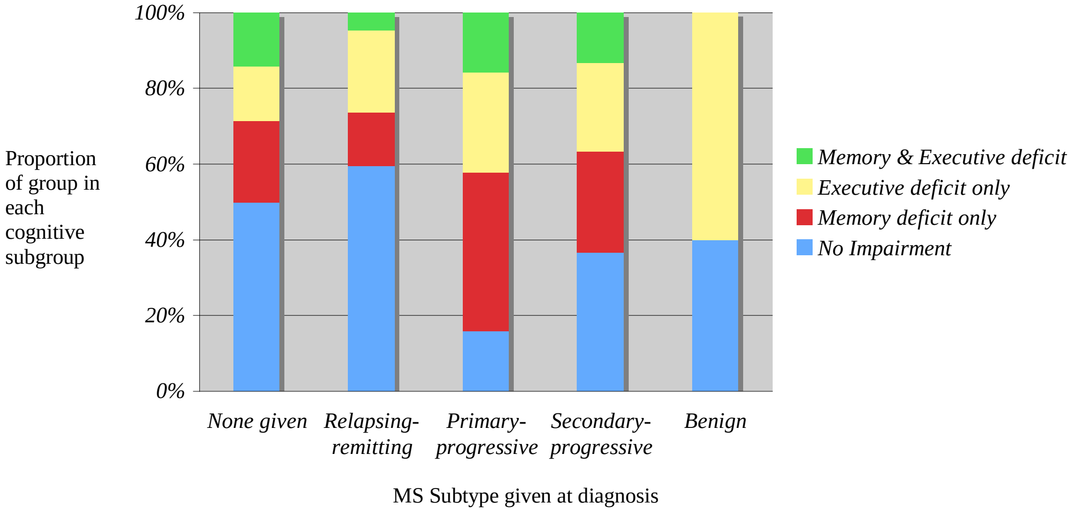


Figure 5.8 Cognitive subgroups organised by MS Subtype. The proportions of each subtype having membership of a cognitive subgroup is indicated

In a one-way analysis of variance, the groups (four levels) created did not differ in age $F(3, 96) = 1.36$, p = 0.260, or in years of schooling $F(3, 96) = 0.766$, p = 0.516. Table 5.17 summarises the proportion of the whole sample in each cognitive subgroup.

Table 5.17: Summary of impairment in the MS sample across domains of cognitive impairment used to investigate heterogeneity in cognitive performance.

	No Impairment	Memory only	Executive Function only	Memory & Executive Function Impairment
Percent of total sample in each impairment group	38%	25%	26%	11%

In line with one proposal in the study, SDMT performance was compared across cognitive performance subgroups (4 levels). ANOVA results suggested statistical differences between the groups on mean SDMT score $F(3, 96) = 6.493$, $p < 0.001$. Post hoc analyses suggested that the *No Impairment* subgroup had better performance on the SDMT than the *memory impaired only* group (p = 0.001), the *executive function & memory* impaired group (p = 0.038) and a difference approaching significance with the *executive impaired only* group (p = 0.074).

Metamemory Measures & Cognitive Subgroups.

When examined according to the cognitive subgroups discussed above, the MS sample showed a pattern of memory self-efficacy (i.e. MFQ responses) indicating more problems (lower scores) being reported by those with either executive deficits alone or executive & memory deficits. While all groups, as well as the sample as a whole gave lower efficacy ratings than the norm groups on the MFQ scales, *Executive only* and *Memory & Executive* groups gave the lowest efficacy ratings. Table 5.18 summarises the increasing memory complaint (lower scores) on the General frequency of Forgetting scale with increasing cognitive impairment, sample size for this comparison is n=96 because of incomplete completion of the scale.

Analysis of variance by cognitive subgroup (4 levels) for 10-item GFF scale did not support reliable differences between the cognitive sub-groups in the sample, though visual inspection suggests increasing differences when compared to the norms, for more cognitively impaired groups.

Table 5.18: Summary of two unimpaired samples and MS sample cognitive sub-groups for the General Frequency of Forgetting scale. n = number of participants in each group; *SD* = standard deviation. For the MS group sample is n = 96.

	10-Item General Frequency of Forgetting scale		
	n	Mean	*SD*
Sample 1	565	4.7	*0.9*
Sample 2	483	4.7	*1.0*
No Impairment	38	4.27	*1.02*
Memory only	25	4.09	*1.18*
Executive only	24	3.83	*.92*
Memory & Executive impairment	9	3.53	*1.07*

Delayed JoL (Predicted minus Actual performance) on AVLT Trial VII.

A one-way analysis of variance compared the mean difference between predicted and actual performance for each of the four cognitive ability subgroups, and suggested that there was a reliable difference between the means of the groups $F(3, 96) = 13.442$, $p < 0.001$. A Tukey post hoc test confirmed that groups of participants without memory impairment (*No Impairment* and *Executive only* impairment) were different from those with memory impairment (*Memory only* and *Memory & Executive Function* impairment); all p-values < 0.01. As shown in figure 5.9 the *No Impairment* and *Executive Function only* groups had greater differences between predicted and actual performance, compared to either of the groups with memory impairment Why memory impaired sub-groups might apparently perform more accurately on this judgment is addressed in the context of structural model results in Chapter 7.

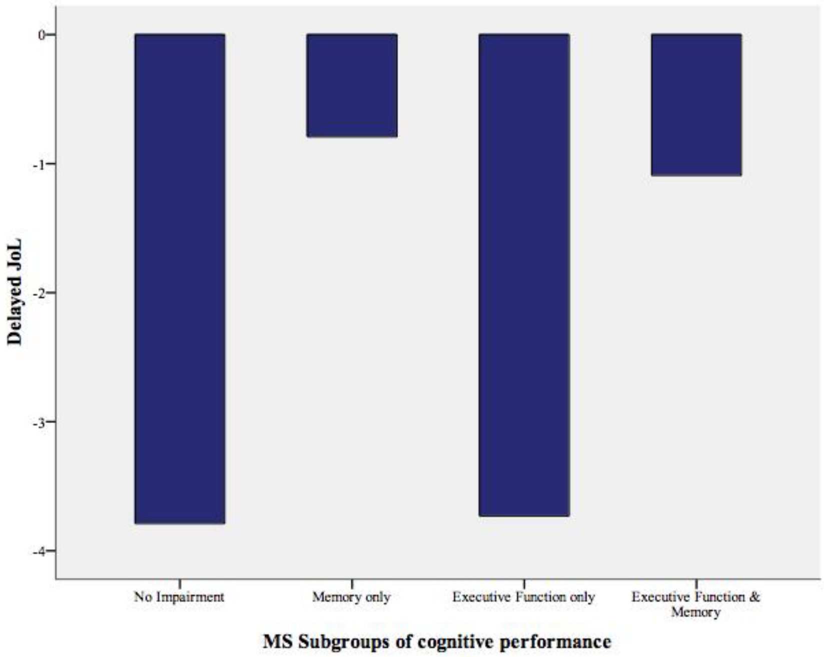


Figure 5.9. Mean discrepancy in *delayed* Judgment of Learning (JoL) organised by cognitive performance subgroup. Delayed JoL is the discrepancy between predicted and Actual recall performance on the delayed recall trial of the Auditory Verbal learning Test. Scores nearer 0 indicate more accurate JoL

Retrospective Confidence Judgment.

On a one-way ANOVA, none of the cognitive performance subgroups (four levels) differed in mean retrospective confidence gamma score; F (3,96) = 0.953, p = 0.418). Generally the results were skewed as most participants had high levels of relative accuracy, on this task.

For absolute accuracy in Retrospective Confidence Judgment, there was evidence of differences between the cognitive performance subgroups' scores. In applying high confidence postdictions to recall, there was a group effect of cognitive performance. One-way ANOVA with post hoc Tukey test of cognitive group (four levels) on % High confidence and correct, suggested that the *Memory impaired only* group had reliably lower scoring than the other three groups F(3,96) = 7.696, p < 0.001. Visual inspection of figure 5.10 suggests that performance of the *Memory only* group was more variable than the others, and that exclusion of a number of outlying scores would might indicate that both *Memory impaired only* and *Memory and Executive Function impaired* cognitive groups would differ from the *No Impairment* and *Executive impaired only* group. This lends some support an interpretation of memory impairment being related to RCJ accuracy, a suggestion that will be investigated in later analyses of structural models. There were no differences between cognitive subgroups for inaccuracy in calibration, that is, mean % Low Confidence correct.

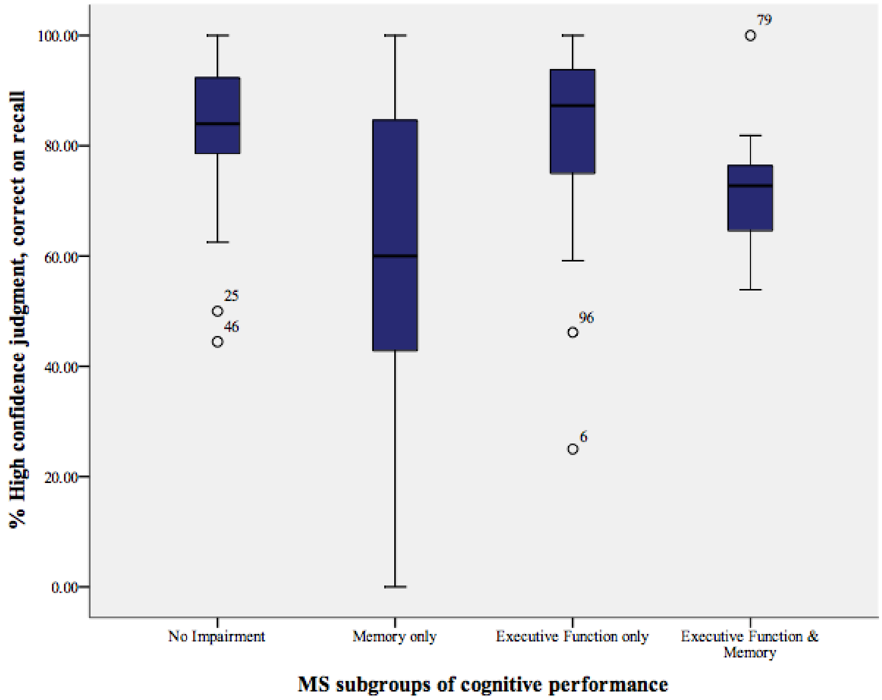


Figure 5.10. Retrospective Confidence Judgment absolute accuracy for each of the cognitive performance subgroups. The figure indicates that participants with memory impairment were more variable in performance compared to others and were statistically lower in the percent of High Confidence rated items that were in fact correct.

Feeling of Knowing gamma correlation

None of the cognitive subtype groups (four levels) differed in mean FoK gamma scoring on one-way analysis of variance $F(3, 96) = 1.366$, $p = 0.258$.

On the % High FoK correct and % Low FoK correct ANOVA for comparison of the cognitive subgroups (4 levels) also indicated no differences in mean % High FoK correct (F (3, 94) = 1,271, $p = 0.292$) or in mean % Low FoK correct (F (3,88) = 0.725, $p = 0.441$).

Metamemory measures and Mood

The Beck Depression Inventory categorises self-reported depression symptoms into 4 categories: None, Mild, Moderate and Severe. These groups were of unequal size as presented earlier. On a one-way ANOVA, only one comparison suggested mood group differences (4 levels) on task-based metamemory measures. Those in the *None/minimal* depression category scored were reliably ($p = 0.001$) less inaccurate on the FoK task than those within the *Mild* depression group $F(3,88) = 5.847$, $p = 0.001$. Visual inspection of figure 5.11 suggests a potential trend towards those being in the *Minimal/None* depression group being less likely to give a low FoK rating to a subsequently correct recognition-based sentence completion.

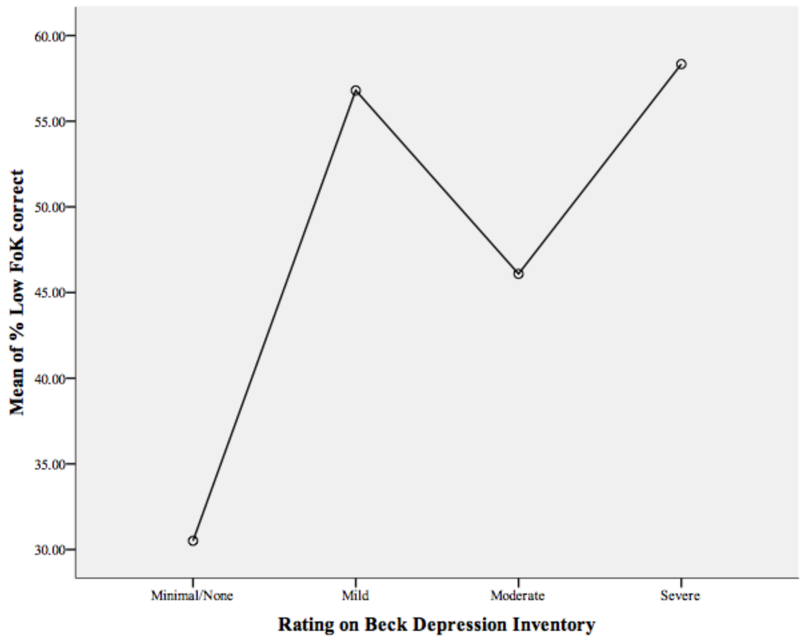


Figure 5.11. Mean percent of Low FoK ratings subsequently recognised for each of the four depression symptom levels on the Beck Depression Inventory 2nd Edition. FoK = Feeling of Knowing.

MFQ scores and Depression symptoms

For the components of interest in this study - GFF-10 and the Forgetting while Reading (FWR), only FWR demonstrated a linear correlation with the total score on the Beck Depression Inventory and suggested that higher scores on the BDI, indicating more reported depression symptomatology was associated with lower scores on the frequency of FWR items, indicating more frequent problems reported in his area (Pearson, $r = -0.25$, $p < .05$, $n=100$).

One way ANOVA, and post hoc analysis of depression symptom category (4 levels) for 33-item GFF, 10-item GFF and Total MFQ scores was carried out and indicated that the *Moderate* group (n = 13) reliably endorsed more memory difficulty that the *Minimal/none* (n = 64) and *Mild* (n = 18) depression symptoms group, but not the *Severe* group (n=5). For Total score on MFQ: $F(3, 89) = 4.443$, $p = 0.006$, for 33-item GFF $F(3, 89) = 3.777$, $p = 0.013$ and for GFF-10 $F(3, 91) = 3.198$, - 0.027). Post hoc Tukey tests also supported no difference between the *Severe* group and any others, despite appearances from visual inspection of mean scoring. Figure 5.12 gives an example, on the GFF-10.

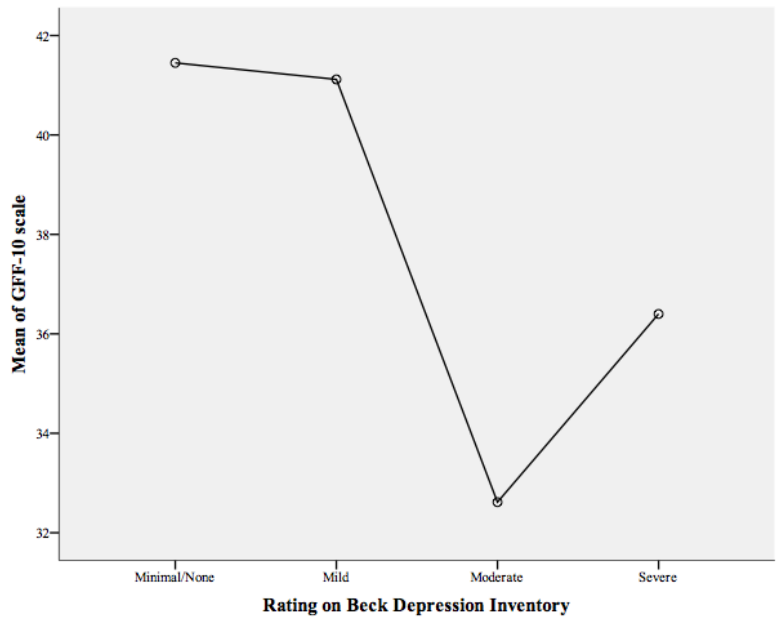


Figure 5.12. Mean score on 10 item General Frequency of Forgetting scale (GFF-10) for each of the four depression symptom levels on the Beck Depression Inventory 2nd Edition. Lower scores on GFF-10 indicate more reported memory difficulties.

5.6. Discussion

This sample of 100 people with Multiple Sclerosis accords well in demographic detail with two recent community dwelling cohorts from the UK, and reflects a number of disease related patterns. The mean age for diagnosis was in the 3rd decade, in line with findings from a number of studies (e.g. Compston & Coles, 2002). Initially high, the proportion people with a relapsing-remitting course declines over time, in conjunction with an increasing proportion of secondary-progressive subtypes. This supports the findings that there may be a change in the disease process for the initially relapsing-remitting group (Weinshenker, et al., 1989; Noseworthy et al., 2000).

Additionally, the frequent diagnosis of primary-progressive MS in later years (5th decade) is reflected in the sample proportions (Weinshenker, et al., 1989). The absence of a benign subtype in the 7th decade may relate to changes in the 'benign' expression of the disease also. There are few very long-term studies following the outcome of people initially diagnosed with benign MS, with which to verify this interpretation (Costello et al., 2008).

The male to female ratio in the sample is not typical of gender-related incidence in the UK, which appears to be about twice as many women as men, possibly with increasing proportions of women. Generally males with MS can have poorer outcomes (Tomassini & Pozzilli, 2009) and so may be harder to recruit to such studies. McCabe, McKern & McDonald (2004) also suggested that men with MS tended to less often employ appropriate coping styles and more often failed to seek support, and so may not be available to recruitment through support organisations. Beatty & Aupperle (2002) suggest than men with MS may be particularly vulnerable to cognitive impairment, which may itself limit participation. Alternatively, and as suggested by Schwartz & Fox (1995), it may be that responders may have time available because they are out of work. Although almost 60% of either sex was retired, just over twice as many of male participants were still in full time employment, compared to women.

In this sample, 38% of the sample had no impairment in memory or executive function. 36% had memory impairment, 26% had executive impairments 11% had impairments in both memory and executive function. 36% of the sample was considered mildly to severely depressed, on a self-report measure of depressive symptoms, but there was no reliable relationship between scoring on the Beck Depression Inventory and cognitive performance subgroup, gender or MS subtype. Information processing, measured by an oral only trial Symbol Digit Modalities Test, was impaired in 36% of the sample and poorer performers on this test were associated with the impaired memory ($p < 0.05$), memory and executive subgroups ($p < 0.05$), but not executive-only subgroup. This finding offers some support for proposals in the general MS literature that these deficits are one way of understanding higher level cognitive deficits in MS (De Luca, et al., 2004; Demaree et al., 1999; Goverover et al., 2007) Letter-Number sequencing performance indicated an impairment level of 23% and Digit Span performance was in line with standardisation samples.

In a global assessment of task-based metamemory measures, there was evidence of sensitivity to task performance and task demands on the Judgment of Learning task. This was evidenced by an underconfidence with practice effect (Koriat, Sheffer & Ma'ayan, 2002); this is where, with repeated exposure to the same learning task, there is a tendency for prediction increase to lag behind performance increases, leading to a discrepancy between predicted and actual performance. Koriat, Sheffer & Ma'ayan (2002) discussed a pattern for this effect, including slight overconfidence in the initial trial - also noted in this study.

The second performance component that may support sensitivity, this time to task difficulty is the difference in the 4 to 1 ratings used in the Feeling of Knowing and Retrospective Confidence Judgments in the study. 73% used all 4 ratings in the RCJ task, compared to 17% in the FoK task. As proposed by Moulin (2002), rank use may be diagnostic of task difficulties in people with Alzheimer's disease. Here, on what was an easier task (RCJ) ranks use was consistently greater than on the FoK task.

Accuracy, measured by gamma correlation, was high on the RCJ task. The RCJ task has traditionally established high levels of accuracy (e.g in traumatic brain injury; Kennedy 2001). Greater variability was noted among those with memory impairment. These judgments are considered to be based on memory experience cues, such as ease of retrieval, fluency of retrieval and perhaps some content cues (Koriat et al., 2008). One possibility is that with memory impairment may come much greater variability in available cues to estimating confidence in a retrieved answer, or that in the face of reduced mnemonic cues (fluency, accessibility), less useful information-based cues (e.g. cue familiarity) may be relied upon, giving wider parameters of accuracy.

Mean Feeling of Knowing gamma was low. One study, which reports a FoK gamma correlation on a similar task, is that of Shimamura and Squire (1986). That study compared a control sample to samples with a range of memory impairments; amnesics, people receiving ECT and people with Korsakoff's syndrome, the latter associated with memory and executive deficits. The findings of FoK performance show controls achieved a mean FoK gamma of 0.70 and amnesics 0.40, both higher than this study's sample. The Korsakoff's group achieved a mean FoK gamma of zero. The authors conclude that this would indicate that more widespread cognitive impairment is probably therefore responsible, rather than relatively circumscribed memory impairment, for success in this task.

In this study, accuracy on the Feeling of Knowing task, may have been limited, in part, by task complexity in the ranking procedure, a feature suggested by a number of authors (Shimamura & Squire, 1986; Beatty & Monson, 1991; Pannu & Kaszniak, 2005). The second is that, being based on an episodic task, it may present more difficulty (Schwartz & Metcalfe, 1994). While one benefit of using a new learning task is in providing equivalence of learning across participants (McKenna & Warrington, 2000), a limitation for the FoK judgment is that magnitude of the FoK relating to the degree of prior learning of the material on which the FoK is based (Nelson & Narens, 1980); here, only a single learning trial was given so likely attenuating the size of the gamma correlation, though some participants did reach gamma correlations of 1.0, and the modal score was 0.50, which compares well to other studies.

FoK calibration, measured by proportion of recognised sentence completions for High and Low FoK judgments was also examined. Results suggest that only the Low FoK, correctly recognised, were associated with other measures. Findings indicate that people not endorsing mood disorder were least likely to give Low Feeling of Knowing ratings to correctly recognised sentence completions. This only reached a statistical difference when compared to the mildly depressed individuals. Given the measure is based on correct recognitions, Mood may be a factor which contributes to the higher number of Low FoKs provided; this effect may be direct or indirect. Direct effect would suggest overly negative appraisals of Feelings of Knowing, and indirect may relate to a high number of 'don't know' responses, which were recorded as low FoK in the study design. People with low mood or more cognitive impairment often use 'don't know' responding more than non-depressed individuals (Foreman et al., 2003). This result will be investigated when mood, memory and executive abilities are considered simultaneously in structural modelling.

In measures of memory self-efficacy, people with MS reported more problems, and considered these more serious when they happened, than non-neurologically impaired people, even though the norm samples against which they were compared were older. An older sample would be typically expected to endorse more memory difficulty than younger samples in part because of implicitly held theories about aging and memory decline (Hertzog, 2002). It may be that a similar disease-related bias also contributes to the samples rating of their own memory on this questionnaire, and this bias may be contributed to by mood disorder. Reported frequency of Forgetting While Reading (FWR) correlated with total score on the Beck Depression Inventory (r = -0.25) indicating more complaints of this memory disturbance with higher depression symptomatology. For the memory Function Questionnaire total score and 10-item General Frequency of Forgetting (GFF-10) scale lowest memory complaint was related to lowest depression report.

These findings are in line with a number of studies of depression and memory self-report in MS (Julian, Merluzzi & Mohr, 2007; Randolph Arnett & Freske, 2004). The finding is further supported by no cognitive subgroup differences between memory self-efficacy ratings, suggesting that cognitive impairment per se is not the primary factor in memory self-report differences. In comparison to neurologically unimpaired samples, all cognitive subgroups (including the no impairment group) reliably rated themselves as having more memory difficulty on the GFF-10, the key measure in this study. Interestingly, GFF-10 scores did correlate with performance on a single trial measure of 60-minute delayed recall, but not with a multi-trial measure of 20-minute delayed recall. These two memory tests may reflect different aspects of experienced memory performance - the single trial task, acquisition or encoding processes and the multi-trial task, forgetting or retention processes. The experience of people with MS is typically of having difficulty with the encoding aspects of memory (Lezak, Howieson & Loring, 2004), rather than forgetting, so single trial tests of this nature may best associate with memory experience as sought in the self-efficacy scale.

A final point of discussion relates to the heterogeneity of performance in the sample, which was quantified by memory and executive performance, and also considered as categories of mood. Neither cognitive performance nor mood category related to MS subtype. Neither did cognitive subgroup consistently relate to any set of variables under study. In contrast to Beatty & Monson (1991) it would seem that categorising participants according this cognitive profile might not assist understanding on metamemory here, at least when a series of parallel statistical comparisons would need to be substantial.

There were no statistical differences between cognitive subgroups in their mean accuracy on the Retrospective Confidence Judgment gamma scoring, though on absolute accuracy (% correct for High and Low confidence), memory impairment was associated with lower accuracy. In contrast to that finding, cognitive subgroups with no memory impairment (*No Impairment* or *Executive only impairment*) had reliably greater underconfidence with practice than the other subgroups. Those participants with memory impairment appeared to be more accurate in their estimation of learning after a 20-minute delay.

Feeling of Knowing gamma was unrelated to any particular subgroup of cognitive function. This is actually in partial agreement with the findings of Beatty & Monson (1991) who demonstrated FoK gamma was different from zero, among controls and the cognitively unimpaired group, but no different from zero in the cognitively impaired sub groupings.

An alternative approach to creating discrete subgroups based on cognitive performance is to examine the cohort as a whole. Exploring the relevance of multiple variables, such as in a regression based approach, may provide a better characterisation of the factors associated with accuracy in a multivariate way. As discussed in this and preceding chapters, the use of well-defined variables is important in this venture - results here suggest that some memory measures indicate relationships and others not. In addition, the relationships of variables are likely to be different in different metamemory measures and interact in different ways. Assessment of sub-group mean performance, while useful as an initial assessment of relationships, may obscure useful and multivariate relationships.

The next analysis will focus on elucidating the relationships between these various findings from the point of view of how performance on the various metamemory measures may be understood based on the cognitive and affect factors examined. This analysis will be carried out through the testing of a number of proposed models to establish if they are acceptable, both theoretically and statistically. The theoretical basis has been explained during the review of the literature and in deriving the objectives for the study. The statistical fit of the data is established next through two processes; first through assessing the validity of the proposed latent variables using confirmatory factor analyses and second through structural model testing. This first step, called the *Measurement Model* focuses on the validity of assumptions about the measures being used so as to underpin the second step, the *Structural Model*, which aims to statistically test the theoretical hypotheses presented in Chapter 1. The next chapter will therefore focus on the Measurement Model and the confirmation of the proposed latent variables - *Memory, Executive Function, Mood, Information Processing* and *Metamemory*.

Chapter 6: The Measurement Model

6.1. Introduction

The measurement model tests assumptions that specific observed variables are congeneric in describing the latent variables of interest. The method for testing these assumptions of factoral relatedness is Confirmatory Factor Analysis (CFA), and the factors established from this analysis are termed *latent variables* in the field of Structural Equation Modeling (SEM: Schumacker & Lomax, 2004). In this chapter the measurement models are tested for each of the proposed latent variables; Memory, Executive Function, Information Processing and Mood. Specification for CFA each model will be set out in advance of testing. Each of the proposed latent variables is created from a number of observed variables, selected and discussed in Chapter 2: *Development of Methods.*

The processes related to model testing have been outlined in the Chapter 3: *Development of Statistical Methods*. In summary, once data has been screened for quality, these steps consist of model *Specification,* testing for model *Identification,* model *Estimation,* assessment of model *Fit,* model *Modification* or *Respecification* and reassessment of *Fit* (Bollen & Long, 1993; Schumacker & Lomax, 2004). While the process is iterative, in this study models tested were confirmatory, rather than exploratory, in that the expectations about factorial structure were specified *a priori* based on a number of published studies.

For each measurement model, the process will be the same - assessing the quality of the data (normality, skew, kurtosis) followed by testing of candidate models. The first measurement model will be of the Beck Depression Inventory, 2nd edition (Beck Steer & Brown, 1996), as this best addresses all stages of the CFA process. Second, will be the measurement model for cognitive factors and finally, the memory efficacy scales. Two measurement models for the Memory Function Questionnaire subscales are investigated - for *Forgetting While Reading* and 10-item *General Frequency of Forgetting scale*.

No measurement model is proposed for task-based metamemory judgments as it has been established that different indices, such as Feeling of Knowing or Retrospective Confidence Judgments tap different monitoring processes, making them unlikely to be factorially related (Dunlosky & Metcalf, 2009).

6.2. Measurement Model for Mood - Beck Depression Inventory, 2^nd edition *(Beck Steer & Brown, 1996).*

The Beck Depression Inventory proposes to represent a latent *Mood* variable, based on 21 indicators of depression, such as changes in sleeping pattern, negative thinking styles and crying (Beck et al., 1961; Beck, Steer & Brown, 1996). Previous factoral studies have proposed one, two and three factor models based on all 21 items or on a smaller number (Beck, Steer & Brown, 1996; Cole et al., 2004; Harris & D'Eon, 2008). Additionally, it has been proposed that some items are differentially sensitive to different factors, depending on the population (Bruce, Polen & Arnett, 2007); an example is the proposal that in people with MS some indicators of depression, such as fatigue, may actually be indicators of neurological disability (Nyenhuis et al., 1995). Many factoral models have been derived from earlier versions of the scale; the more contemporary version used here has been structured to reflect more current dimensions of depression, and considers a longer period of time in its assessment of mood disorder.

6.2.1. Data Screening

Missing Data

There were no missing responses for the BDI.

Normality

Much of the data collected in behavioural research does not accord with univariate or multivariate normality (Micceri, 1989; Finney & DiStefano, 2006; West, Finch & Curran, (1995) discuss the implications and management of non-normality in their treatment of the topic in Hoyle (1995). Two specific issues for the BDI were; the expected lack of normality in an assessment that measures depression in a heterogeneous group and the categorical response format (four options).

Limitations to the process of model fitting for the BDI-II have been recognised elsewhere, in respect of non-normality of data (Osman, et al., 1997; Cole et al., 2004). Other issues include the appropriateness of assumptions about underlying continuity of the data represented by the four response categories (Byrne, 2006), and concern over the small sample size for such a CFA, which generates a large number of to-be-estimated parameters, and therefore degrees of freedom (McIntosh, 2007). Subsequent to the degrees of freedom issue, non-normality of data is also relevant in assessing the criteria for model fit – many goodness-of-fit measures are impacted by degrees of freedom as well as non-normality of data (West, Finch & Curran, 1995). Simulation studies reported by the authors have focused on testing known models with varying degrees of non-normality and different sample sizes, in order to test the robustness of indices of fit, or of some kind of x^2 adjustment for non-normality. One finding outlined is of the x^2 goodness-of-fit test rejecting more than 5% of true models (West, Finch & Curran, 1995).

The combined effect of non-normality, small sample sizes and many free parameters associated with a 21-item scale like the BDI, is to increase the occurrences of 'Heywood cases' (West, Finch & Curran, 1995; Brown 2006:189). These are cases where model solutions cannot be achieved because of negative error variances occurring. Non-normality may also reduce some of the approximate indices of fit, such as the Comparative Fit Index (CFI; West, Finch & Curran, 1995).

A final reason for problems in reaching statistical solutions is poor model specification (Brown, 2006; McIntosh, 2007). Therefore not all problems are therefore data related; some can be theory-related. Ultimately, acceptance or rejection of a theoretically plausible model is based on a number of criteria. The fit indices, parameter values and a range of other statistical judgements are the initial assessment of acceptability. Substantive theory is the other.

In summary, item skew, kurtosis, the presence of outliers can be indicative of both univariate and multivariate non-normality (Field, 2005). In turn, these will impact on computing model solutions, accuracy of fit statistics and stability of solutions. For this reason each was considered in relation to this dataset, prior to analysis taking place. Cole et al., (2004), in validating the structural properties of a similar rating scale, report many similar considerations to those presented here – inadmissible solutions because of negative variance and positively skewed data.

Univariate normality.

Item univariate Skew and Kurtosis, and associated standard errors, were tabulated and are presented in Table 6.1. Critical ratios suggest that data is skewed and kurtotic (values > 2.58 or greater; Field, 2005). Some authors suggest that kurtosis is a more significant concern with non-normality, because of its impact on model estimation (Byrne, 2001; West, Finch & Curran, 1995). The positive direction of the skew is as expected because most respondents did not report significant mood disorder, and therefore scored at the lower end of the scale.

Table 6.1: Univariate and Multivariate distributions of 21 Beck Depression Inventory II items. c.r. = Critical Ratios for Skew and Kurtosis. These are equivalent to Skew and Kurtosis values divided by their standard error, and represents z-scores (Field, 2005: 72). Assuming normality, Skew/Kurtosis have a mean = 0 and a standard error = 1. Critical ratios of > 2.58, plus or minus, indicate a non-normal distribution. Multivariate normality is indicated by Mardia's (Mardia, 1970) coefficient, which also indicates multivariate non-normality when values > 2.58

Variable / Univariate Normality	Skew	c.r.	Kurtosis	c.r.
Loss of Energy	-.014	-.055	-.549	-1.121
Fatigue	.745	3.041	-.127	-.258
Crying	1.726	7.048	2.954	6.029
Agitation	1.802	7.359	3.880	7.920
Loss of Interest	2.013	8.220	4.071	8.310
Indecisive	1.043	4.259	.129	.263
Irritability	1.750	7.144	2.151	4.392
Changes in Sleep	.873	3.566	.112	.229
Changes in Appetite	1.579	6.447	1.834	3.744
Loss of Concentration	.487	1.988	-.748	-1.527
Loss of Interest in Sex	1.310	5.347	.636	1.299
Past Failure	1.258	5.138	1.321	2.697
Pessimism	.906	3.698	.547	1.116
Sadness	1.397	5.703	.703	1.435
Self-Dislike	1.226	5.005	1.532	3.127
Self-Criticalness	1.234	5.039	1.254	2.560
Suicidal Thoughts or Wishes	1.858	7.584	2.624	5.357
Worthless	1.068	4.362	.369	.753
Punishment Feelings	1.803	7.362	1.891	3.860
Guilty Feelings	1.019	4.160	-.029	-.059
Loss of Pleasure	1.114	4.549	.503	1.026
Multivariate Normality			*105.142*	*16.914*

Multivariate Normality

Mardia's coefficient of multivariate kurtosis had a critical ratio of 16.914, well above the 1.96 value (at $p < 0.05$) or 2.58 (at $p < 0.01$) level at which significant departure is defined. Approaches to handling data that is not normally distributed were therefore investigated. The first consideration is whether the data would be expected to be normally distributed in this sample in the first instance (Ullman & Bentler, 2004) Given the non-psychiatric sample used in this study, the skewed result is expected (Cole et al, 2004); further, with low scores indicating lower levels of depression, a positive skew is expected. Data transformations are typically used to greatest effect with such skewed data (Schumacker & Lomax, 2004) but while transforming data towards normality, they may equally be being transformed away from the true population distribution (Hayduk, 2001)

For untransformed data, a number of methods are available for adjusting for the non-normality in calculating the model fit statistics, specifically the x^2 statistic. Multivariate non-normality typically will impact on the x^2 statistic, on parameter standard errors and some fit indices, though not necessarily the parameter value itself (Hutchinson & Olmos, 1998; Finney & DiStefano, 2006). One mechanism for adjusting the x^2 value for the non-normality of the data is a bootstrap process (Efron, 1979; Bollen and Stine, 1992; Fouladi, 1998). Bootstrap procedures involve random resampling of some of the available dataset, replacing the sampled data points with data that relates to the overall characteristics of that dataset itself. This generates many alternative datasets and these datasets can then be used to test the same model and generate adjusted x^2 estimates by averaging the derived sample statistics (Finney & DiStefano 2006; Blunch, 2008). This would be a bootstrap process related to the x^2 goodness-of-fit statistic to account for non-normality in the original sample (Bollen & Stine, 1993; Byrne, 2001).

The Bollen-Stine bootstrap procedure (Bollen and Stine, 1992; Bollen Stine 1993; Finney & DiStefano 2006; Arbuckle 2007) can be used specifically with more modest sample sizes (Finney & DiStefano, 2006). Alternative bootstrapping procedures, such as the Satorra-Bentler correction (Satorra and Bentler, 1994; Harris, and D'Eon, 2008; Vanheule et al., 2008) can be used, but the recommended sample size is ideally greater than 200 cases (Nevitt & Hancock, 2001). One CFA study of the BDI-II (Vanheule et al., 2008), points to the use of this Satorra-Bentler bootstrap only in a large sample of people.

Some authors discuss the limited evidence at the n = 100 level for bootstrapping, making recommendation difficult (Nevitt & Hancock, 2001; Kline, 2005). Instead, especially for categorical data such as here, some suggest a process called parcelling (Hull, Tedlie & Lehn, 1995). Parcelling is an approach where a number of items on a scale are summed and treated as a single continuous indicator so that more normal estimation methods may give more stable results, by increasing the number of possible scores. The issue with parcelling, especially with a tool where the factor structure seems to differ in different populations (as the BDI-II does; Beck Steer & Brown, 1996), is deciding, a prioi, which items to parcel (Bandalos & Finney, 2001). In general, it is considered a somewhat controversial topic because it requires known information about dimensionality of a scale (Kline, 2005). Given the number of different models of the BDI-II derived from differing samples, such clear dimensionality cannot be assumed. Parcelling was therefore not used, and the small sample size meant that use of the Satorra-Bentler or Bollen-Stine bootstrap adjustment to the x^2 value has only limited support as an approach to management of non-normality. Transformation of the data therefore remained an alternative option.

While one proposed limitation of transforming data such as the BDI-II responses is that the original metric, and therefore some interpretability can be lost. However, results presented in Chapter 5 allow for understanding of the sample performance on this scale. The transformation was considered appropriate from the point of view of generating a more stable model, and from the perspective that various rotations were used in generating the original model that is tested here (Beck, Steer & Brown, 1996). On balance therefore, a square-root transformation was considered to present a beneficial impact on normalising the positively skewed data. The results of the transformation, in terms of univariate and multivariate normality are presented in table 6.2, and better concord with most heuristics about acceptable levels of normality.

Table 6.2: Univariate and Multivariate distribution of square-root (sq) transformed 21 Beck Depression Inventory II items. c.r. = Critical Ratios for Skew and Kurtosis. These are equivalent to Skew and Kurtosis values divided by their standard error, and represents z-scores (Field, 2005: 72). Assuming normality, Skew/Kurtosis have a mean = 0 and a standard error = 1. Critical ratios of > 2.58, plus or minus, indicate a non-normal distribution. Multivariate normality is indicated by Mardia's (Mardia, 1970) coefficient, which also indicates multivariate non-normality when values > 2.58

Variable / Univariate Normality	Skew	c.r.	Kurtosis	c.r.
Loss of Energy	.868	3.543	-.839	-1.712
Fatigue	.009	.038	-1.621	-3.308
Crying	.704	2.874	-1.160	-2.367
Agitation	-.071	-.292	-1.727	-3.525
Loss of Interest	.749	3.059	-.973	-1.987
Indecisive	1.326	5.412	-.030	-.062
Irritability	-.062	-.255	-1.435	-2.929
Changes in Sleep	-1.108	-4.522	.190	.388
Changes in Appetite	.464	1.895	-1.496	-3.053
Loss of Concentration	.406	1.656	-1.494	-3.049
Loss of Interest in Sex	1.194	4.876	-.155	-.317
Past Failure	.761	3.108	-.934	-1.907
Pessimism	1.448	5.912	.300	.612
Sadness	.451	1.841	-1.433	-2.925
Self-Dislike	.513	2.093	-1.448	-2.955
Self-Criticalness	1.268	5.178	-.020	-.040
Suicidal Thoughts or Wishes	.573	2.338	-1.468	-2.997
Worthless	.454	1.855	-1.465	-2.991
Punishment Feelings	.404	1.651	-1.422	-2.903
Guilty Feelings	-.001	-.004	-1.538	-3.140
Loss of Pleasure	1.191	4.862	-.499	-1.018
Multivariate Normality			27.903	4.489

Outliers

Figure 6.1 of individual box-plots was created for the untransformed *z*-scores of each of the 21 items on the scale, and demonstrated a number of outliers for 18 of the 21 items. This figure also underlines the positive skew in the scoring, with most people scoring low and outliers being higher scorers. Extreme values occurred where all but a small number of participants scored anything other than 0 on the scale (0-3) on items S*uicidal Thoughts or Wishes* and more *Irritability* than usual. Figure 6.2 presents the same assessment after square-root transformation; the dispersion of item-by-item responses suggests that this process reduced the number of outlying scores.

Multivariate outliers were sought through the information provided by AMOS about the Mahalanobis Distance (d^2) of scores. None of the observations in the transformed data appear to be 'improbably far from the centroid' (Arbuckle, 2007:240).

Measurement Scale

The issues relating to measurement scale for this instrument have been considered in Chapter 3 and suggest that it is unlikely to present significant cost, especially where data are or approach a normal distribution. Having considered the issues related to the data; normality, outliers, measurement scale, and having addressed them to a pragmatic extent, the square root transformed data was used for all models.

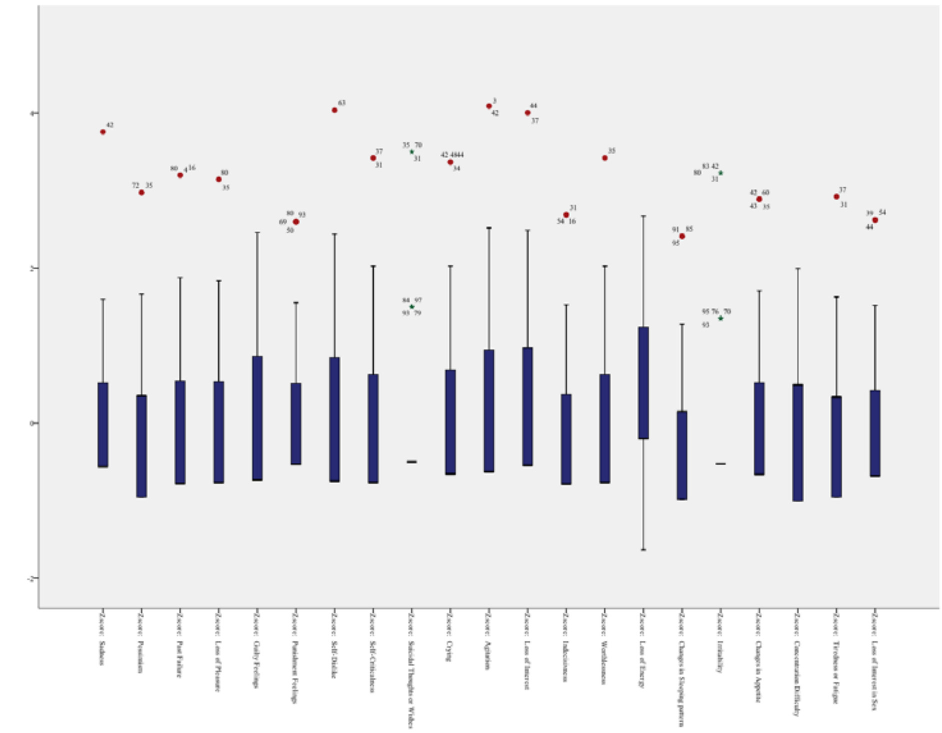


Figure 6.1. Univariate outlier detection using z scores of 21 BDI-II items (n=100). Outlier labels refer to case numbers indicated in red, and extreme values those in green.

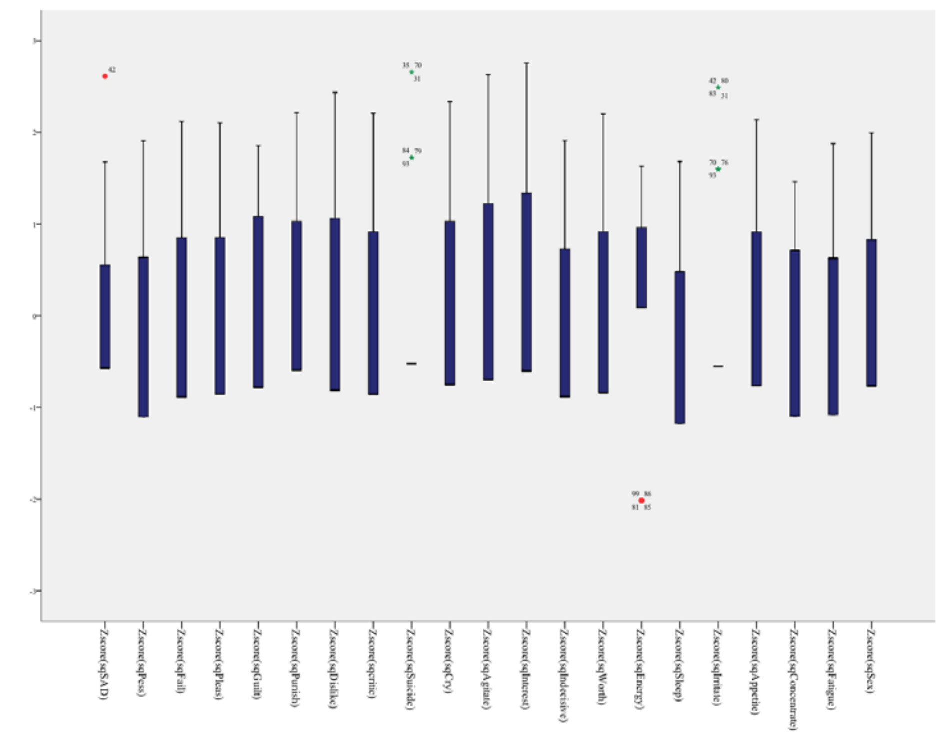


Figure 6.2. Univariate outlier detection using z scores of square root transformed data for the 21 BDI-II items (n=100). Outlier labels refer to case numbers indicated in red, and extreme values those in green.

6.2.2. Model Specification.

Two groups of models were specified; first, models based on published factor structure in the test manual (Beck, Steer & Brown, 1996), with a single-factor model for comparison. Thereafter, testing of factor models based on three clinical samples, each of which might be expected to relate to the sample under study - people with chronic pain, treatment seeking substance abusers, and head-injured people. The reasons for this are discussed in considering theoretical specification of these models.

A single factor model was tested first. It contained all 21 items of the BDI-II. The single factor model is based on a number of studies that suggest that there is a single second-order factor structure to the BDI-II (Gullion & Rush, 1998; Cole, et al., 2004). This was considered a default model against which multi-factor models were assessed. The primary multi-factor models included a two-factor model based on the exploratory factor analyses published in the manual for the BDI-II (Beck, Steer & Brown, 1996) of 500 psychiatric outpatients, and a two-factor model based on exploratory factor analysis of 120 undergraduate students, also presented in the test manual.

Theoretical Specification of Models

Many dimensionality assessments of the BDI-II tend to be taken from the North American undergraduate population (e.g., Beck Steer & Brown, 1996; Osman et al, 1997; Dozois, Dobson & Ahnberg, 1998; Whisman, Perez and Ramel, 2000; Storch, Roberti & Roth, 2004), rather than clinical samples. One reason for this is that in testing and selecting the best fitting model for the BDI-II, the ratio of sample size to degrees of freedom (df) in the models requires large samples. The large sample size requirement for model testing is a challenge in health sciences, but perhaps less so in an undergraduate population. These undergraduate samples tend not to be chronically (medically or psychiatrically) unwell and tend to be young (e.g. Osman et al, 1997 mean age = 19.02 years; Stock, Roberti & Roth 2004 mean age = 20.52 years). This is a limitation in terms of confirming models; in the main they have been developed on a young healthy, non-clinical, sample and not the range of populations for which many such measures have been developed.

In samples of people with ‘fairly minor medical conditions’ (Viljoen et al., 2003:289), more serious medical conditions (Thombs, Zeigelstein, Beck & Pilote, 2008) and in primary care medical attendees (Arnau, Meagher, Norris & Bramson (2001), a two-factor model including *Somatic/Affective* & *Cognitive* factors has been proposed, or confirmed to be as good other two or three factor models (Thombs, Ziegelstein, Beck & Pilote, 2008). All three studies supported the presence of a single second-order depression factor also.

The picture is complicated however when considering people for whom symptoms of depression and somatic or psychosocial factors might be confounded, such as in MS (Nyenhuis et. al., 1995), in drug withdrawal (Seignourel, Green & Schmitz 2007), chronic pain (Harris & D'Eon 2008) or traumatic brain injury (Rowland, Lam & Leahy, 2005). A number of factorial studies of the BDI-II have been carried out which were deemed useful to consider. They include substance abusers, where withdrawal might implicate especially somatic aspects of depression (Buckley, Parker & Heggie 2001), people with chronic pain and people with traumatic brain injury potentially reflecting cognitive, affective and psychosocial results similar to an MS population. The dataset from the MS sample under study was therefore also submitted to these models to assess their potential as better accounts of self-reported depression in MS, than those derived from a younger sample or from people with psychiatric illness.

Statistical Specification

In specifying models, correlated errors (residuals) were allowed if considered theoretically plausible; there is some debate about whether this is statistically appropriate (Schumacker & Lomax 2004). Some suggest that allowing error terms to correlate might indicate some redundancy in inventory items, (e.g. items about fatigue and lack of energy) and is thus appropriate in analysis of tools where redundancy might occur (Byrne, 2001:93; Harris & D'Eon, 2008). In this sample of people with MS where *fatigue* and *loss of energy* might be confounded, it would therefore be acceptable. In the initial models, no error covariances were included, but were considered as part of the modification and re-specification process. Error term covariances were kept to a minimum and were selected based on modification indices once considered theoretically acceptable. Figure 6.3. summarises the models. Each will be outlined prior to testing.

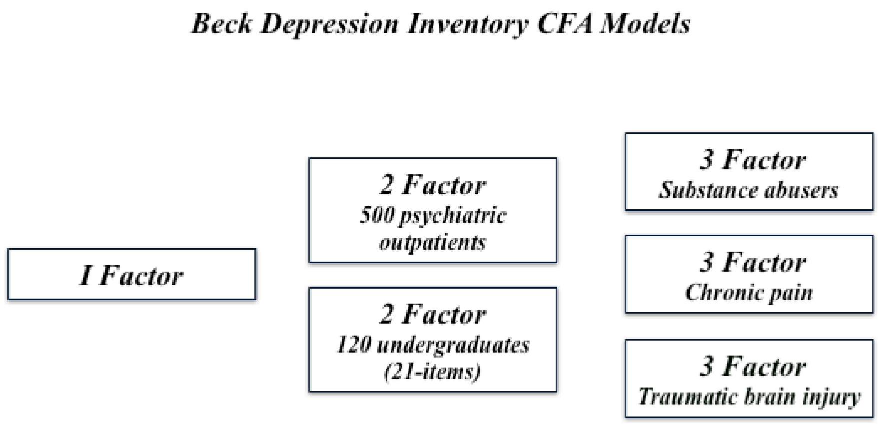


Figure 6.3: Summary of planned Beck Depression Inventory (BDI) Confirmatory Factor Analysis models.

6.2.3. Model Testing

Model Identification.

This identification stage, including other components of identification (e.g. the rank condition: Schumacker & Lomax, 2004:173), are checked as part of the initial AMOS output, and was found to be identified in all models tested.

Model Estimation.

All models were tested using Generalised Least Squares (GLS) estimation in AMOS with transformed data used. Sample size for each model was 100 participants.

Model Fit.

The concept of model fit, and fit indices, was discussed at length in *Development of Methods* where the selected indices of fit were presented. Selected fit indices and other assessments of model acceptability are summarised in Table 6.3, with the acceptable values where appropriate indicated in Table 6.4. All statistical judgments were set a cut off for alpha of < 0.05, unless otherwise stated.

Table 6.3: Fit indices and other assessments of model acceptability. x^2 = Chi Square value; x^2 p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. Interfactor correlations are relevant to the interpretation of Confirmatory Factor Analyses with values > 0.80 suggesting that factors may be too closely relate to be discriminated.

Specific indices of model fit	**Other Assessments**
x^2 & p-value	Standardized Residual Matrix values generally < 2.58
CMIN/df	Interfactor correlations in 2+ factor models < 0.80
GFI	Squared Multiple correlation *(r^2)* values > 0.20
CFI	Parameter sign & significance
RMSEA: 90% CI & p-value	
SRMR	

Table 6.4: Summary of recommended values for model fit indices. x^2 = Chi Square value; x^2p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual

x^2 p-value	**CMIN/df**	**GFI**	**CFI**	**RMSEA**	**SRMR**
p > 0.05	< 2.0	> 0.90	> 0.90	90% CI upper limit of 0.08; pclose > 0.50	< 0.10

6.2.4. Model I: 1-factor model of BDI-II

The model is presented in figure 6.4. Model fit indices were: x^2 = 233.653, df. = 189 p = 0.015, x^2/df = 1.236, GFI = 0.775, CFI = 0.345, RMSEA = 0.049 (90% CI 0.023 - 0.068; p = 0.523). SRMR= 0.1547. Regression weights were all significant at $p < 0.05$ level and positive in value; Standardised Residual Covariances again were generally < 2.58. Finally, item r^2 values for 20 of 21ietms were > 2.0 (*Changes in Sleep* r^2 = 0.184).

Modification indices suggested a number error covariance of which only *Sadness & Pessimism* and *Loss of Pleasure & Crying* were considered theoretically plausible, so were allowed to covary. The model was then retested for fit. Only *Sadness & Pessimism* had an expected positive error covariance and the value was significantly different from zero (p = 0.018; r = 0.310). Respecified model fit statistics were x^2 = 228.103, df. = 188, p = 0.024, x^2/df = 1.213, GFI = 0.781, CFI = 0.412, RMSEA = 0.046 (90% CI 0.019 - 0.067; pclose = 0.578). Regression weights were all significant at $p < 0.05$ level and positive in value, Standardised Residual Covariances were generally < 2.58 in absolute values and for item r^2 values all but one was > 2.0 (*Changes in Sleep* r^2 = 0.189).

The x^2 p-value in this model was < 0.05, which should lead to outright rejection of the model, according to Barrett (2007). Given that some other indices present acceptable results, including the x^2/df value (1.213), it will be retained for comparison with the other candidates from the rest of the model testing process. Table 6.5 summarises fit indices for both the initial and modified 1-factor model, and Appendix D the statistical data for the final model.

Table 6.5. 1-Factor BDI: Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. n/a = factor correlation not applicable because only a single factor.

Model 1	x^2	**df**	**p**	x^2**/df**	**GFI**	**CFI**	**RMSEA**	**90% CI RMSEA**	**RMSEA p value**	**SRMR**	**Factor correlation**
Initial Model	233.653	189	0.015	1.236	0.775	0.345	0.049	0.023 -0.068	0.523	0.1547	n/a
Modified Model	228.103	188	0.024	1.213	0.781	0.412	0.046	0.018- 0.067	0.595	0.1520	n/a

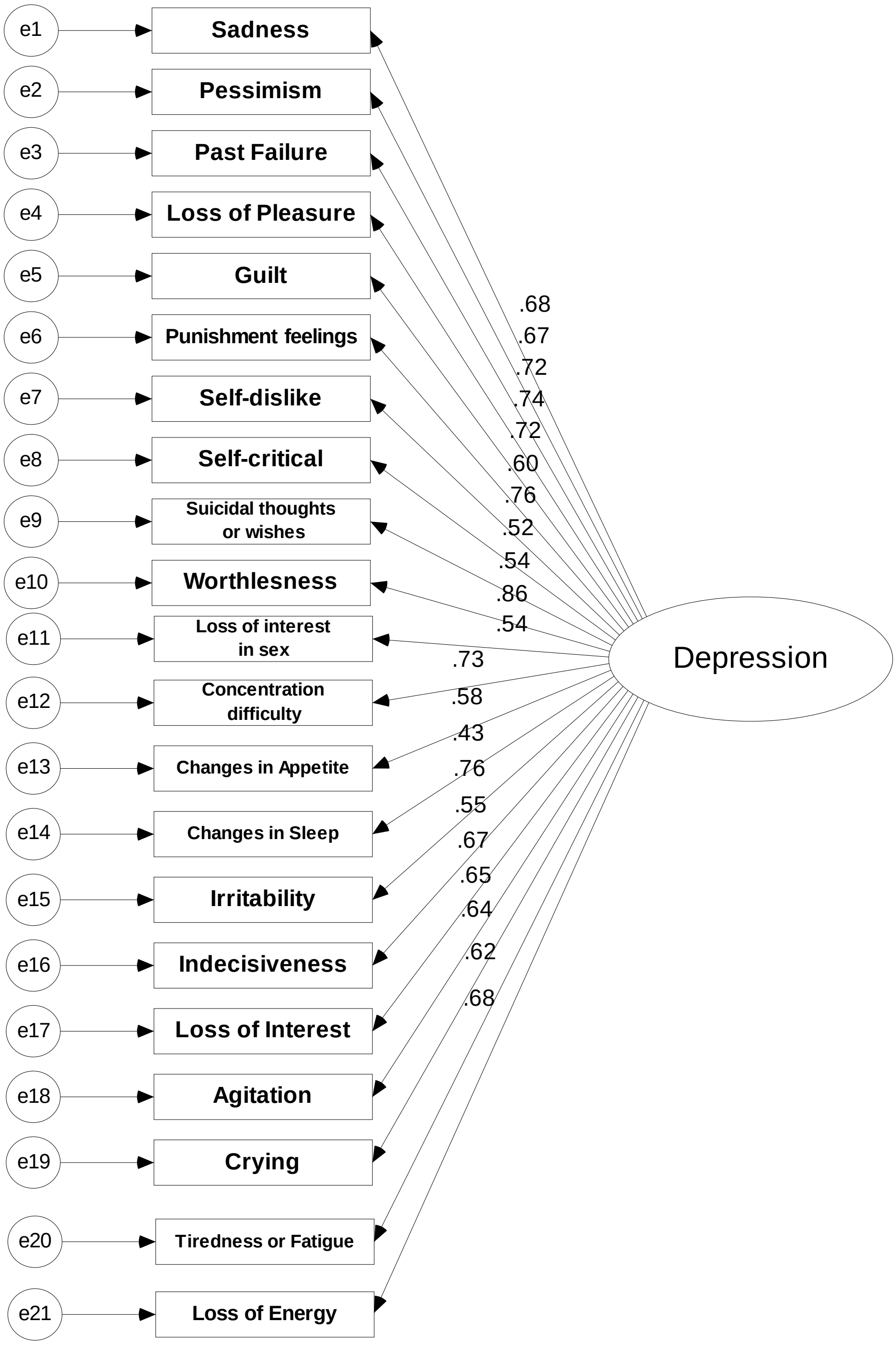


Figure 6.4. I Factor Model of the Beck Depression Inventory. Parameter estimates are the factor loadings for each of the 21 observed indicators. These are standardised regression weights. Model fit statistics were x^2 = 228.103, df. = 188, p = 0.024, x^2/df = 1.213, GFI = 0.781, CFI = 0.412, RMSEA = 0.046

6.2.5. Model II: 2-factor model of BDI (500 psychiatric outpatients). Factor 1: Cognitive, Factor 2: Somatic/Affective.

The results of initial model fitting were: x^2 = 229.131, df =188 p = 0.022, x^2/df = 1.219, GFI = 0.780, CFI = 0.397, RMSEA = 0.047 (90% CI 0.019 - 0.067; pclose = 0.578), and SRMR = 0.1259. Parameter estimates did not meet the criteria for significance in a number of instances, with 10 of the 20 estimated parameters being non-significant. Inter-factor intercorrelation was also non-significant (p = 0.432; r = 0.624). The Standardised Residual Matrix presented large number of values > 2.58, suggesting model-implied and data-implied covariance structures differed considerably. From the r^2 values 7 of 21 were < 0.20. Overall this suggested a poor fit of the model based on both fit indices and individual parameter assessment.

None of the modification indices were reasonable from a theoretical perspective, so were not pursued. Issues related to the non-significant interfactor correlation, parameter estimates and significant model-implied versus data-implied differences, underlined in the matrix of Standardised Residuals. The fit indices are presented in table 6.6 below, but should be interpreted with caution, as many other criteria for a theoretically acceptable model: regression weights that are significant, significance and size of the inter-factor correlation, were not met. As a result, it is proposed that this is a mis-specified model for this sample. Of interest, this result presents a good example of conflicting findings between global fit and component lack of fit. The full statistical outputs are presented in Appendix E.

Table 6.6. Model II. 2-Factor BDI: 'Cognitive' and 'Somatic/Affective'. Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. n.s. = non-significant

Model II	x^2	df	p	x^2/df	GFI	CFI	RMSEA	90% CI RMSEA	RMSEA p value	SRMR	Factor correlation
Final Model	229.131	188	0.022	1.219	0.780	0.397	0.047	0.019 - 0.067	0.578	0.1259	0.624 n.s.

6.2.6. Model III: 2-factor BDI based on 120 undergraduates.
Factor 1: Cognitive/Affective, Factor 2: Somatic.

In the original development of this model, Beck, Steer & Brown (1996) removed the two items with lowest factor loadings (*Pessimism* and *Loss of Interest in Sex*). As this judgement was based on the results of an exploratory factor analysis, and in line with others (Osman et al, 1997), *Pessimism* and *Loss of Interest in Sex* were included with the Cognitive/Affective factor here because higher regression coefficients were associated with each in the original Beck Steer and Brown (1996) analysis.

Results were as follows: x^2 = 220.751, df = 188, p = 0.051; x^2/df =1.174, GFI = 0.788, CFI = 0.520, RMSEA = 0.042 (0.00 - 0.063; pclose = 0.712), and SRMR = 0.1331. Correlation of factors was 0.766 ($p < 0.05$) and all parameters were both positive in value and significant. In the matrix of Standardised Residual covariances *Loss of Interest in Sex*, *Crying* and *Agitation* again seemed to be the least well described items, generating values > 2.58. r^2 values for 20 of the 21 items were > 0.20, with *Sleep* again < 0.20.

The highest modification index value (8.228) was for an error covariance between *Past Failure* and *Self Dislike*. The estimated value of the covariance was (standardised r = -0.46) which was not meaningful because of the negative relationship, so was not applied. Only a *Sadness* and *Pessimism* residual covariance was theoretically acceptable and resulted in a value of r = 0.30 (p = 0.021). Overall model fit statistics were: x^2 = 215.477, df = 187, p = 0.075; x^2/df = 1.152 GFI = 0.793, CFI = 0.582, RMSEA = 0.039 (90% CI 0.00 - 0.061; pclose = 0.770), and SRMR = 0.1308. Correlation of factors was r = 0.769. The matrix of standardised residual covariances indicated *Loss of Interest in Sex*, *Crying* and *Agitation* remained problematic. Standardised (regression) parameter loadings were all significant and positive in value, and factor correlation between 'Cognitive/Affective' and 'Somatic' was r =.796. r^2 values for all 21 items, except *Crying* were > .20. This model was therefore considered acceptable in terms of fit assessment, and was therefore retained for later model comparison. Table 6.7 summarises the fit indices for the initial and modified model, and Appendix F the statistical data for the final model. This CFA model is presented in Figure 6.5.

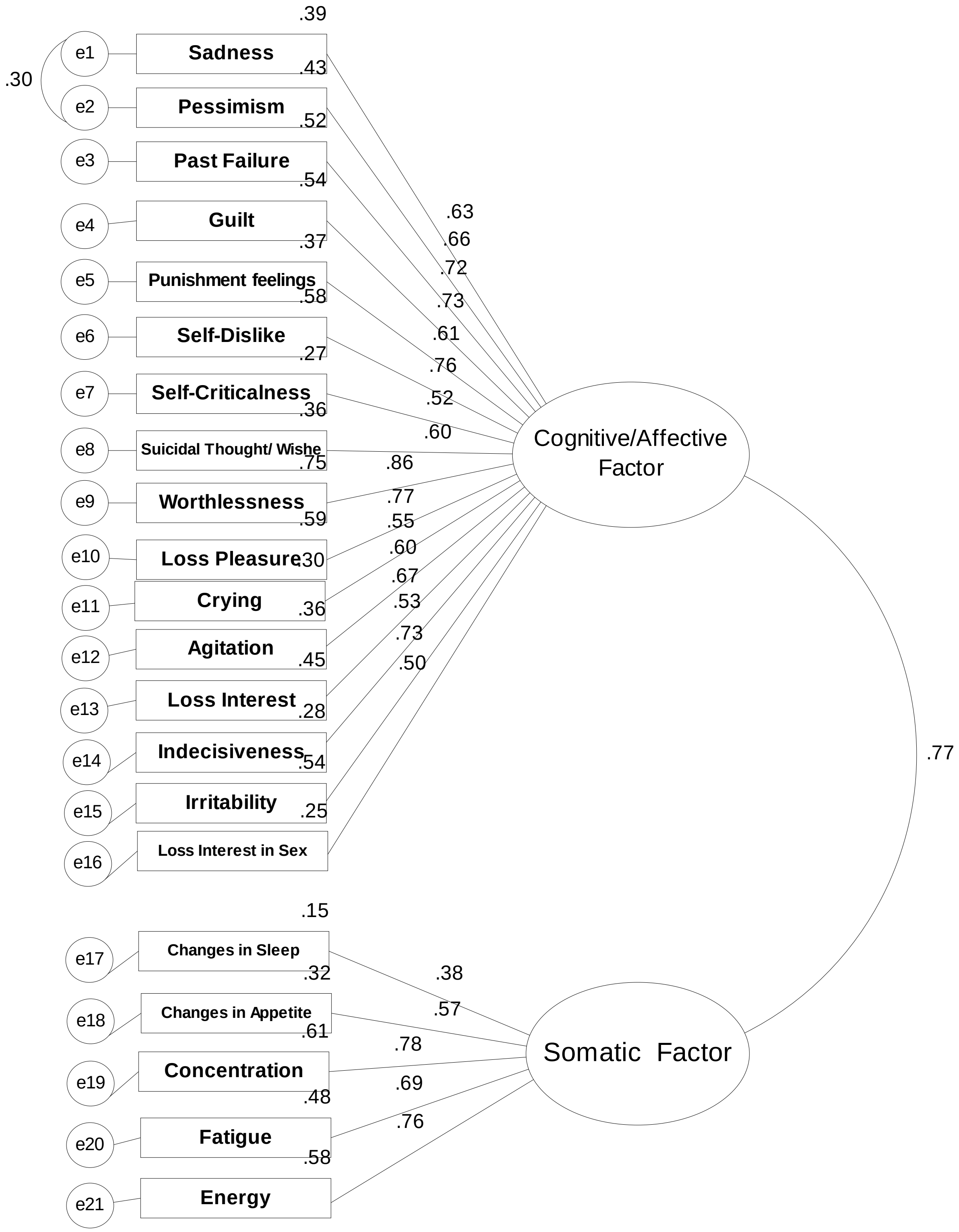


Figure 6.5. Beck, Steer & Brown's (1996) Two-factor Confirmatory Factor model of the Beck Depression Inventory, based on a sample of 120 undergraduates. Factor 1: Cognitive/Affective, Factor 2: Somatic. Final fit statistics were x^2 = 215.477, df = 187, p = 0.075; x^2/df = 1.152 GFI = 0.793, CFI = 0.582, RMSEA = 0.039.Error covariance between two observed items is modelled

Table 6.7. 2-Factor 21 item BDI: 'Cognitive/Affective' and 'Somatic' Model based on 120 college students, from Beck Steer & Brown, 1996. Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. Factor correlation is $p < 0.05$

Model III	x^2	**df**	**p**	x^2**/df**	**GFI**	**CFI**	**RMSEA**	**90% CI RMSEA**	**RMSEA p value**	**SRMR**	**Factor correlation**
Initial Model	220.751	188	0.051	1.174	0.788	0.520	0.042	0.000 - 0.063	0.712	0.1331	0.766
Modified Model	215.477	187	0.075	1.152	0.793	0.582	0.039	0.000 - 0.061	0.770	0.1308	0.769

6.2.7. Model IV: 3-factor BDI-II model on a sample of 'treatment-seeking substance abusers'. (Buckley, Parker & Heggie 2001; Seignourel, Green & Schmitz 2007).

The model estimation gave a non-positive definite (NPD) covariance matrix error, so the model solution is considered inadmissible. This is likely because of multicollinearity between BDI items. The reported correlations were: Cognitive Factor & Affective Factor = 0.96, Affective Factor & Somatic Factor = 0.92, Cognitive Factor & Somatic Factor = 0.75.

6.2.8. Model V: 3-factor model of BDI-II in people with chronic pain (Harris and D'Eon, 2008).

A study of people with chronic pain was also selected because this is a population where somatic and depressive symptoms might concur (Harris & D'Eon 2008). One difference might be in the level of severity of endorsed depression symptoms; the MS sample might be more heterogeneous and people with chronic pain tend to more often report more severe symptoms of depression (Harris & D'Eon, 2008). A final note was that the chronic pain sample was of an average age of 43.25 years, more in line with the age of participants in this study. The model has 3 factors; Factor 1: Negative Attitude, Factor 2: Performance Difficulty and Factor 3: Somatic elements. Figure 6.6 presents this model. In the original study, inter-factor correlations were high with r's = 0.84, 0.98, 0.95 in women and r's = 0.85, 0.98 and 0.96 in men. Not surprisingly therefore, a single second-order factor model was also deemed an acceptable model. Here, the 3-factor model was tested.

Results were x^2 = 216.952, df = 186 p = 0.060, x^2/df = 1.166. GFI = 0.791, CFI = 0.546, RMSEA = 0.041 (90% CI = 0.00 - 0.062; pclose = 0.732) and SRMR = 0.1393. Parameter estimates and residual covariances were all acceptable. Modification indices suggested a number of potentially acceptable within-factor residual covariances between *Past Failure* & *Self Dislike*, and *Guilt* & *Self Criticalness*. Only *Guilt* & *Self-Criticalness* was positive in value. The addition of the error covariance between *Guilt* & *Self-Criticalness* led to a re-estimation of model fit. Main fit indices were: x^2 = 211.551, df = 185 p = 0.088, x^2/df = 1.144. GFI = 0.796, CFI = 0.610 RMSEA = 0.038 (90% CI = 0.00 - 0.06; pclose = 0.790) and SRMR = 0.1381. Statistical fit and interfactor correlations for iterations of this model are presented in Table 6.8. The full output is available in Appendix G. The high inter-factor correlation in this model argues for a reduced number of factors - either 1 or 2 factors being potentially a better theoretical fit and more parsimonious.

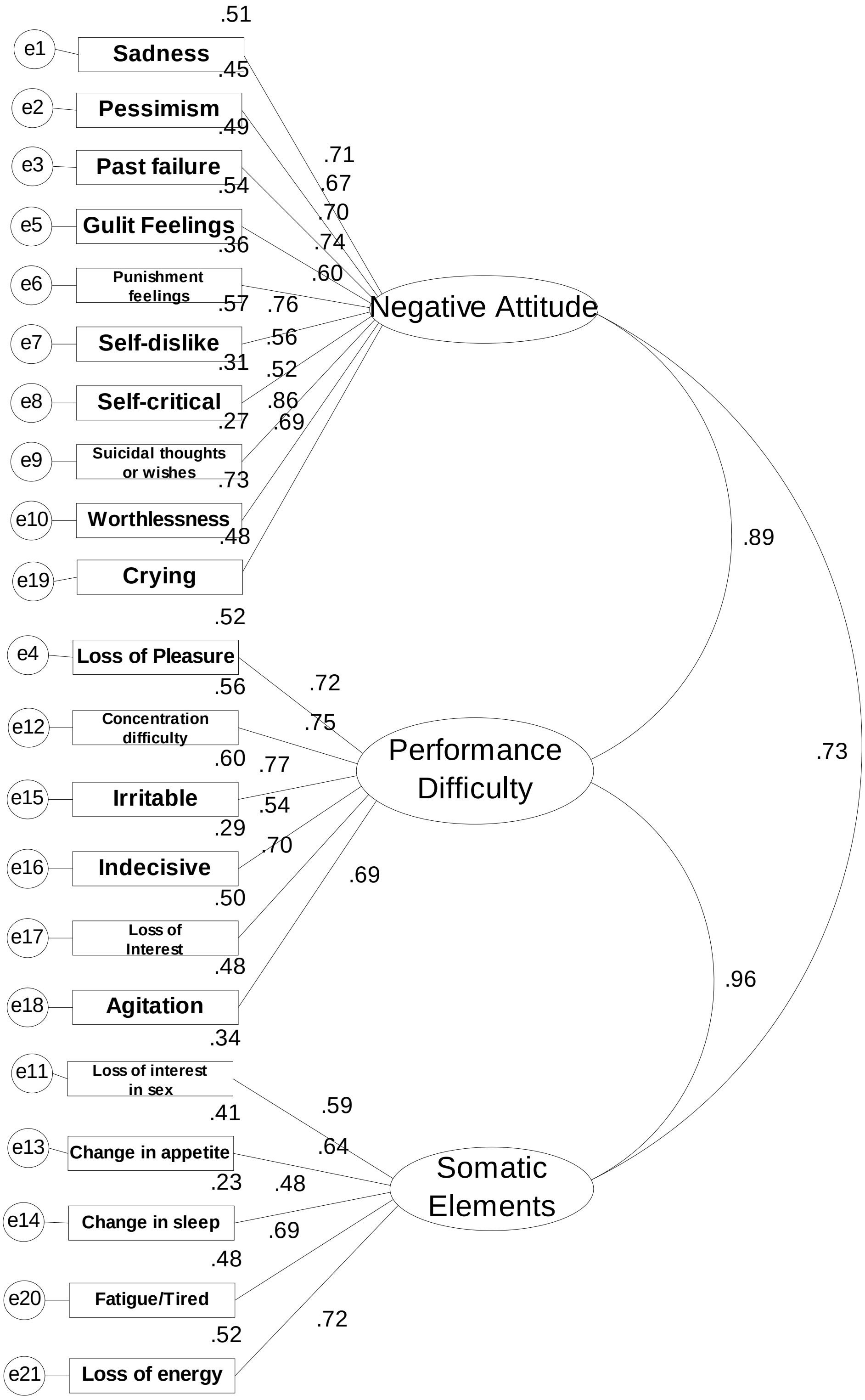


Figure 6.6.The Harris and D'Eon (2008) three-factor Confirmatory Factor model of the Beck Depression Inventory, based on a sample of people with chronic pain. Factor 1: Negative attitude, Factor 2: Performance difficulty & Factor 3: Somatic elements. Model fit for the initial model was $x^2 = 216.952$, df = 186 p = 0.060, $x^2/df = 1.166$. GFI = 0.791, CFI = 0.546, RMSEA = 0.041

Table 6.8. 3-Factor Harris & D'Eon (2008): Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. All interfactor correlations are P <0.05, but at r > 0.80, some are too large to be considered as separable factors.

Model V	x^2	**df**	**p**	x^2**/df**	**GFI**	**CFI**	**RMSEA**	**90% CI RMSEA**	**RMSEA p value**	**SRMR**	**Factor correlation**
Initial model	216.952	186	0.060	1.166	0.791	0.546	0.041	0.00 - 0.062	0.732	0.1393	0.960, 0.890, 0.734
Final model	211.551	185	0.088	1.144	0.796	0.610	0.038	0.00 - 0.06	0.790	0.1381	0.958, 0.893, 0.737

6.2.9. Model VI: 3-factor BDI model - Traumatic Brain Injury (Rowland, Lam and Leahy, 2005).

Finally, a sample of people with brain injury was investigated (Rowland, Lam and Leahy, 2005). This paper was considered because it implicated time since injury as a relevant factor in increasing depression, suggesting that the BDI was sensitive to a psychosocial contribution to measured mood disorder, which has been discussed also in respect of MS-related depression (Minden 2000).

This model generated a non-positive definite matrix, with the 3 interfactor correlations at r's = 0.997, 1.109 and 1.011. It seems likely that the factor description is not sufficiently well specified. While a correlation value of > 1.0 is not typically found, the AMOS software may estimate a correlation > 1.0 for a number of reasons including linear dependence between items, that is collinearity (Brown, 2006). Many would suggest that the appearance of Heywood cases in CFA could also relate to poor model specification (Schumacker & Lomax, 2004; Kline, 2006; Blunch 2008).

6.2.10. Discussion of Results.

Generally, high interfactor correlations (> 0.80) argued for smaller numbers of factors and many published studies also accept that there may be a single second-order 'self-reported depression symptoms' factor. From some perspectives single factor models may be of limited clinical utility. In a sample of people with MS for example it may be clinically useful to be able to separate different factors. The ability to discern whether one set of items is more affected by a disease process, is a better indicator of depression or is more responsive to intervention offers greater sensitivity (Nyehuis et al., 1995; Randolph et al., 2000). It might also be of use in determining the relationships of the disease itself to reporting affective as opposed to somatic symptoms of depression, given the potential for somatic symptoms to reflect aspects of the disease - fatigue, loss of energy, sleep or appetite changes (Nyenhuis et. al., 1995). In conjunction with analyses of the normative samples of the BDI (Beck, Steer & Brown, 1996) a two-factor structure does seem applicable.

Conclusions about best fitting models were necessarily tentative with an approximate ratio of participants to each parameter (n=100; BDI-II items = 21; parameters to be estimated =

42-47). Many of the measures of model fit are strongly related to the degrees of freedom (df) within a model (Steiger, 1990; Hox, 2002, Cole et al., 2004) and with 21 items in the models, df were between 185 and 188. Thus, measures of fit, based in part of df, such as GFI and CFI (Cole et al., 2004; Arbuckle, 2007), were expected to be impacted; in all models these were below desired levels (typically < 0.80, rather than > 0.90 for GFI). Measures less impacted by sample size or df in the model, such as the RMSEA generally were not expected to be impacted, and all values were < 0.05, and had desired p value of > 0.50. In assessing model fit therefore, three assessments were made - Exclusion of inadmissible models, appraisal of *Fit Indices* (notably the x^2 and associated p-value), and between-model comparison.

Typically modification indices were used to improve model fit, in this case based on correlating residual error terms. This decreases the df and is also a plausible modification based on similarity in some questions – e.g. *Sadness* & *Pessimism* or *Sadness* & *Crying* questions might share the some redundancy in the scale. In applying modifications to models, a light touch was maintained both because of the philosophy behind confirmatory approaches and because with the aim of creating factor scores based on the best fitting CFA model, these modifications would be lost. Factor assignment would be reflected in the factor score, but residual covariance would not.

Inadmissible solutions

The literature appears clear that a single dimension of ‘depression symptoms’ is reflected in the BDI-II. This presents the possibility of encountering issues related to collinearity. As a result, the first set of judgements related to models for which collinearity was possibly a problem leading to their rejection, or inadmissibility. These typically involved generation of non-positive definite matrices, inter-factor correlations that were too high (r > 0.80; Maruyama, 1998) or regression weights that were non-significant or have the wrong sign (here minus values). Other reasons for initial rejection of models were where there were unacceptable differences between the model-implied and data-implied covariances/correlations. These latter differences became evident when the matrix of Standardised Residual Covariances was inspected and many values were > 2.58 (Joreskog & Sorbom1998; Arbuckle, 2007). In these cases it is proposed that the discrepancy between the model and data is unacceptable.

Model V: Harris and D'Eon (2008)

With interfactor correlations of 0.96, 0.89 & 0.76 in the Harris & D'Eon (2008) model, 3-factor dimensionality was discounted in favour of a more likely 2- or single-factor model. As the aim of the modelling process here is to test a measurement model rather than explore factor structure, and because of the high interfactor correlations, it is more realistic to reject these models as not being optimal in fit. The Harris & D'Eon (2008) model was therefore not selected for further evaluation.

Model IV: Rowland, Lam and Leahy (2005) & Model V: Seignourel, Green & Schmitz, (2007); *Buckley, Parker & Heggie (2001)*

In a similar vein, the Rowland, Lam and Leahy (2005) model of people with traumatic brain injury, and the Buckley et al., model (Buckley, Parker & Heggie, 2001; Seignourel, Green & Schmitz, 2007) of treatment-seeking substance abusers were excluded because of their generation of non-positive definite (NPD) matrices. The basis of the SEM is in matrix algebra; for mathematical operations to take place on these matrices, a matrix needs to be a 'positive definite' matrix. NPD matrices arise when this is not the case, and the reasons for this frequently relate to high levels of linear dependence between items on the BDI (Schumacker & Lomax, 2004:48; i.e. between the 21 items on the BDI-II), that may not have been accounted for in the correlations or 'used-up' by the latent variable structure (Blunch, 2008:93), suggesting model mis-specification for this sample. The inter-factor correlations of near and above 1.0 also support a specification error in the model, possibly related to poor item-factor assignment not applicable to this sample (Schumacker & Lomax, 2004)

Model II: Beck, Steer & Brown, 1996 - Cognitive & Somatic/Affective factors

One other model worthy of consideration in respect of high correlations between factors was the Beck model (Model II: Beck, Steer & Brown, 1996) based on a sample of 500 people in an outpatient psychiatric setting. This model, with Cognitive and Non-Cognitive (Somatic-Affective) factors, generated a non-significant correlation of $r = 0.624$. Additionally, about half of the regression weights in the model were also non significant. As the matrix of standardized residuals contained many large residual values, this model is probably misspecified by poor item-to-factor assignment (Maruyama, 1998:67). In summary, this is likely to be a misspecified model, given non-significant regression weights, a high number of large values in the Standardised Residual Matrix, high standard errors for variance estimates and a non-significant correlation between the latent variables, despite a value of $r > 0.60$. One cause of this may relate to the conflation of affective and somatic items. In neurological samples (as opposed to psychiatric samples) somatic items might be seen to be congeneric, because of their relationship to the disease process, rather than anything to do with mood disorder (Nyenhuis, et al., 1995).

Admissible solutions

Assessment of a 2-factor model, this one based on a sample of 120 university undergraduates, and the single factor model remains warranted. In addition the relatively consistent findings of a second-order single factor of ‘self-reported depression’ (Steer et al., 1999:126), a model presenting two correlated factors - *Cognitive/Affective* and *Somatic* was also considered as an acceptable fit. Results of fitting are presented in Figure 6.8. Using only the x^2 and associated p value, the two-factor model was the only of the two that met requirements for global fit. Given the outlined limitations of sample size to estimated parameters, consideration of other indices demonstrated little difference between these models; x^2/df, RMSEA or GFI.

Model comparison was therefore carried out using a x^2 difference test to confirm whether there is a statistical decrement in fit between the models. Using the initial version of each model, a x^2 test of difference (1-factor x^2 *minus* 2-factor x^2) gives $x^2 = 12.902$, with $df = 1$, supporting a reliable difference between the models at $p < 0.001$. This supports the 2-factor model, because it demonstrates that fit is reduced in the single factor. By effectively setting the correlation of the 2 factor model to = 1.0 (in doing so making a single factor, more constrained, model), the x^2 difference test result says there is ‘no validity in imposing this constraint’ (Raykov & Marcoulides, 2006:102).

A final comparison investigates if allowing correlated errors between *Sadness* and *Pessimism* supports them being related in a conceptual sense. The correlated error terms effectively assesses whether these items share error variance, exclusive of their relationship to a *Cognitive/Affective* depression factor. This again is tested by a x^2 test of difference (2-factor no correlated error *versus* 2-factor no correlated error) gives x^2 =

5.274, with *df* = 1, having a difference of $p < 0.05$ supporting the addition of a free-to-be-estimated parameter, the residual correlation.

Statistical comparison supports the 2-factor model with a *Cognitive/Affective* and *Somatic* factor, in line with the findings of Beck, Steer and Brown (1996) and supported by Dozois, Dobson & Ahnberg, (1998), which included a cross validation sample and a gender comparison that supported this as a generalisable factor structure.

6.3. Measurement Model for Cognitive Items

Introduction

In this measurement model three cognitive latent variables were initially proposed – *Executive Function*, *Memory* and *Information Processing*. Measurement of each has been discussed in Chapter 2, *Development of Methods*.

6.3.1. Data Screening

Missing data

Two missing scores from partial test completion were generated, both related to difficulties with the tasks; the data is therefore treated as *non-ignorable*. Remaining missing data (for 2 additional participants) for these models were treated as *missing completely at random* (MCAR), in that it related to experimenter error for 2 instances of the Brixton test. Regression-based single-imputation was carried out based on the each model tested, to preserve characteristics of the dataset and because of the low instances of missingness.

Univariate Normality.

Absolute value for skew and kurtosis, and associated critical ratios are presented for each measure in Table 6.9. Absolute values for most are acceptable, in line with the previously discussed guidelines. The Hayling test performance was, according to critical ratios for kurtosis, non-normal, however the dependence between errors and time (i.e. a speed accuracy trade off) may explain this pattern.

Table 6.9: Univariate and Multivariate distributions of Cognitive tests. c.r. = Critical Ratios for Skew and Kurtosis. These are equivalent to Skew and Kurtosis values divided by their standard error, and represents z-scores (Field, 2005: 72). Assuming normality, Skew/Kurtosis have a mean = 0 and a standard error = 1. Critical ratios of > 2.58, plus or minus, indicate a non-normal distribution. Multivariate normality is indicated by Mardia's (Mardia, 1970) coefficient, which also indicates multivariate non-normality when values > 2.58

Variable / Univariate Normality	Skew	c.r.	Kurtosis	c.r.
Letter Number Sequencing (LNS)	.241	.982	.640	1.307
Symbol Digit Modalities Test (SDMT)	.085	.349	-.495	-1.011
Digit Span Task (Digits)	.614	2.507	-.084	-.172
Sentence Memory Test (SMT)	.350	1.428	-.061	-.125
Auditory Verbal Learning Test (AVLT)	-.768	-3.134	.364	.706
Hayling Errors A-type (A Errors)	1.746	7.127	2.924	5.968
Hayling Errors B-Type (B Errors)	.777	3.171	-.297	-.607
Hayling Time (HayBminusA)	1.709	6.977	3.097	6.321
Brixton Test Errors (BrixErr)	.230	.938	-.582	-1.189
Multivariate Normality			*7.587*	*2.696*

Multivariate Normality

Mardia's coefficient for Multivariate normality (MVN) critical ratio was 2.696, indicating only a small departure from MVN.

6.3.2. Model Specification

Theoretical Specification of model.

Selection of measures was based on the *a prioi* model of likely factor structure, the observed variables for which have been discussed earlier. The *a prioi* 3-Factor CFA is made up of: Factor 1: *Information Processing* (SDMT, LNS and Digits) Factor 2: *Memory* (AVLT7 and SMT recall) and Factor 3: *Executive Function* (Hayling A errors, Hayling B errors, Hayling Time and Brixton Errors). As discussed in Chapter 4: *Methods* the Hayling time score was calculated to exclude the possible confound of processing speed by subtracting simple initiation time from response time, in the difficult sentence completion task. Executive scores were reversed so that higher scores for all observed variables indicate better performance.

6.3.3. Model Testing

6.3.4. Model I - 3 Factor Model: Information Processing, Memory & Executive Function.
In this model, the Letter Number Sequencing task (LNS), Symbol Digits Modalities Test (SDMT) and Digit span test (Digits) indicate the latent variable of Information Processing, Executive Function is indicated by 4 Hayling & Brixton measures and Memory by the Auditory Verbal Learning Test and Sentence Memory Test.

Results from initial model fit were x^2=33.772, df = 24 p = 0.089, x^2/df = 1.407. GFI = 0.924, CFI = 0.816 RMSEA = 0.064 (90% CI= 0.00 - 0.111; pclose = 0.30) and SRMR = 0.0899. Global indices of fit were fairly encouraging and regression weights were all within expected size and were positive. Correlations between factors were the expected sign - positive between Executive Function and Memory, Executive Function and Information Processing and between Memory and Information Processing. The size of correlations was acceptable, and all were significant. Full statistical output is available in Appendix H, and the model is shown in Figure 6.7.

Of interest in regard to model fit was a low CFI score (0.816), low factor loading for the Brixton test (r^2 = 0.123), and values in the Standardised Residual covariance matrix associated with this item > 2.0. Unlike the models based on the Beck Depression Inventory, normality was within acceptable ranges for cognitive items, and the measures are perhaps more factorially validated, so the low (near significant) x^2 p value in conjunction with CFI, residual matrix and r^2 factor loading was worthy of assessment. While acceptable model fit is suggested, given the discussion in Chapter 1 about factoral distinctiveness of executive function, working memory and information processing, some model improvements were considered. This process focused on theoretical considerations more than statistical.

One issue in investigating this result relates to three of the four measures of executive function being from the same task - the Hayling Sentence Completion task. The measures taken from the task are *Time*, number of *category A errors* and number of *category B errors*. Scoring on these is not independent - e.g. a response time score is likely to relate to difficulty with the task and difficulties on this test might also lead to errors of word generation, of either A or B type. Effectively, participants may be slow to answer in order to maintain a low error rate, or fast but errorful to increase speed. A and B-type errors are also mutually exclusive categories so may imply an association with each other. The relationship between these indicators might equally well be to do with the fact that they are generated from the same task performance. Such a common methods effect could explain the small amount of Brixton performance explained by this latent factor, largely indicated by the three Hayling task measures. Given that the normed data of the Hayling & Brixton tests (Burgess & Shallice, 1997) suggest they are not strongly correlated, a feature of executive tests already discussed, the Brixton test was kept as a contributor to the latent variable. Maintaining the Brixton measure is an issue of theoretical specification given the proposed diversity of executive function.

Aside from the loading of the Brixton tests as an executive function item, the relationship of the LNS task to executive function is worthy of consideration. The latter was considered first as there was reason to suspect, from the literature, that it could load on the executive function factor.

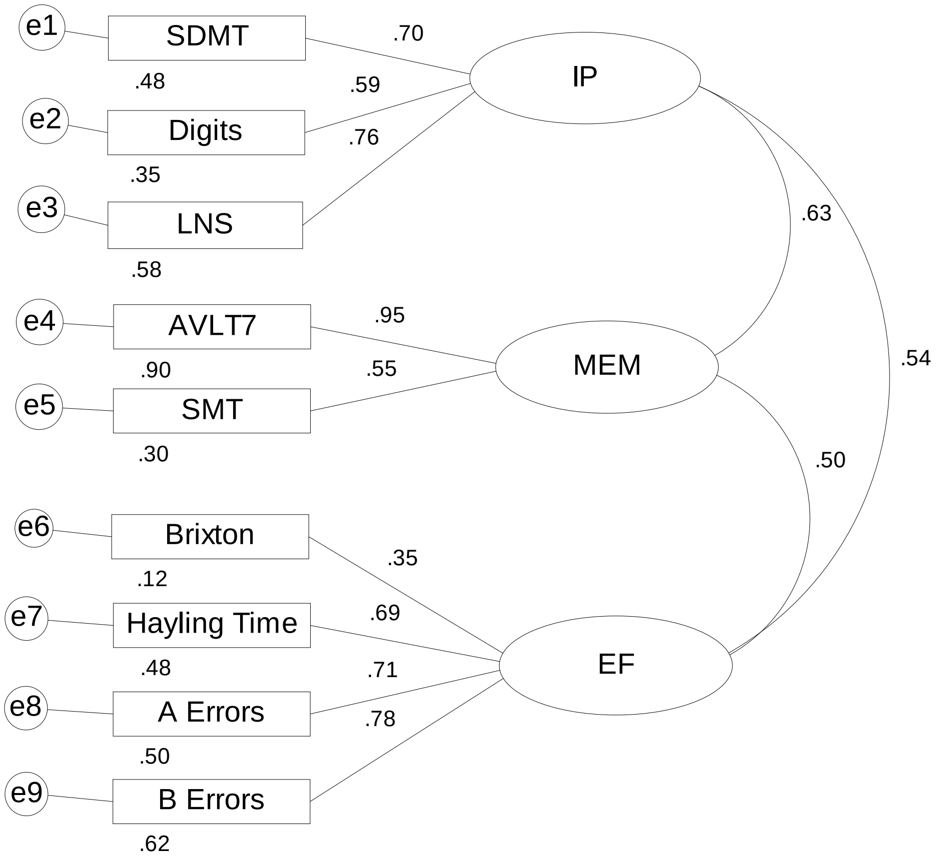


Figure 6.7. The a priori Cognitive items' CFA Model 1. IP=Information Processing, MEM= Memory, EF= Executive Function. SDMT=Symbol Digit Modalities Test, Digits=Digit Span Task, LNS=Letter Number Sequencing, AVLT7 = Auditory Verbal Learning Test (trial 7), SMT = Sentence Memory Test, A Errors/ B Errors - A type/ B type Errors Hayling Sentence Completion test. x^2=33.772, df = 24 p = 0.089, x^2/df = 1.407. GFI = 0.924, CFI = 0.816 RMSEA = 0.064. Interfactor correlations are all significant at $p < 0.05$.

6.3.5. Model II - 3 Factor Model – Information Processing Capacity, Delayed Recall and Executive Function (including LNS).

This model generated a non-positive definite matrix and an inter-factor correlation of 1.07 between Information Processing and Executive Function. This might suggest that the LNS item shares variance with both the information processing and the Executive Function dimensions, but equally could indicate that one of the other items in either factor is leading to the problem (Drew, Starkey & Isler, 2009). The single element, which might underlie each factor description, is the SDMT, proposed as a mediating factor (Information Processing) in the structural models, and in cognitive performance in MS more generally (DeLuca, et al, 2004).

One of the key aims of developing a structural model was to investigate the mediating effect of an information-processing factor on the relationship of other cognitive operations (i.e. Memory and Executive Function) to metamemory. Two possible models could fulfil this function. The first alternative is a factor description in which information processing speed (SDMT) and complex processing or working memory abilities are related, but not congeneric. This decision relates to the discussion in Chapter 1 about lack of clarity in describing the processing deficit in MS. Therefore a four-factor model is proposed where the SDMT is a single indicator of information processing speed, the Digits and LNS indicate a 'Working Memory factor' with Memory and Executive Function remaining the other two factors. A second option, accepting the LNS might cause difficulties because of shared variance across two factors, was allowing it to cross-load on both executive function and information processing, though this would undermine the aim of modelling it and other information processing items as a mediators in later models. This latter approach was not considered warranted so a 4-factor model was specified with correlated factors and no item-factor cross loadings.

6.3.6. Model III. 4-Factor Model - Information Processing, Working Memory, Memory & Executive Function.

In order to meet model identification requirements for the 4-factor CFA, error variance had to be explicitly modelled in the case of single indicators. This is recommended where the input is a covariance matrix, as it is here, by setting the path to 1.0 and the measurement error variance to a fixed non-zero value (Joreskog & Sorbom, 1999; Schumacker & Lomax, 2004; Blunch 2008) and the non-zero value should be, according to Bollen (1989):

*(1 - reliability) * varSDMT*

where varSDMT is the observed variance of the SDMT, r is the estimate of internal or split-half reliability; for SDMT this is reliability is not reported in the literature, so a value of 0.80 was given; variance of sample performance was calculated to be var = 153.904. Measurement error was therefore given a value of 30.7808 for the unstandardised model, with a factor path loading of 1.0. The 4 factors for this model were therefore; Information Processing (SDMT), Working Memory (Digits, LNS), Memory (AVLT and SMT) and Executive Function (3 Hayling & 1 Brixton measures)

Results of model fitting were: x^2 = 24.370, df = 22 p = 0.328, x^2/df = 1.108. GFI = 0.945, CFI = 0.955 RMSEA = 0.033 (90% CI= 0.00 - 0.093; pclose = 0.617) and SRMR = 0.0697. All regression paths were positive and significant, interfactor correlations were significant and as expected, positive in sign. r^2 values were all acceptable, with Brixton Errors the smallest, but improved, value at r^2 = 0.20. The matrix of standardised residuals was acceptable with all values < 2.58. All three factors correlated to a reliable extent. The full statistical output is presented in Appendix I.

The results of the a priori 3-factor model and the 4-factor model were compared statistically, given that both might be considered to fit. This was carried out by a x^2 difference test. The comparisons gave a x^2 value of 9.402 for the 2 degrees of freedom difference between the models (model 1 *minus* model 2), which suggests a reliable improvement on model 1 fit statistics (p = 0.01). In conjunction with good values of CFI and in the standardised residual matrix, this supports a 4-factor account of the data from the sample.

6.3.7. Discussion of Results of Cognitive CFA Models

In all, three models were tested with two models being good candidates. Both of the solvable models presented a non-significant chi-square (x^2) p-value indicating, at first pass, an acceptable global fit (Barrett, 2007). However when consideration was given to components of the models, Model I may have had some problems with specification leading to some elements of poor fit. This was the *a priori* model, but it was rejected because of apparent theoretical mis-specification, in favour of a model that seemed to address suboptimal distinctions between factors. Findings here therefore suggest that an additional *Working Memory* factor be considered, separate to both Information Processing and Executive Function.

Model III had 4 factors; Information Processing, Working Memory, Memory and Executive Function. This four-factor model offers a distinct, and theoretically founded, differentiation of processing speed and working memory abilities. This finding is in line with a range of published investigations suggesting limited coherence in the concept of a single information-processing factor in Multiple Sclerosis (Chiaravalloti, et al., 2003; Drew, Starkey & Isler, 2009), that speed and attentional demands in the context of task complexity may reflect different factors, rather than a single coherent one.

Generally, the evidence in MS suggests that information processing speed, rather than working memory, is the primary deficit (Demaree et al., 1999; DeLuca, et al, 2004; Lengenfelder et al., 2006; Diamond et al., 2008), though they are related, and may be more so for some MS subtypes than others (Archibald & Fisk, 2000; Drew, Starkey & Isler, 2009). In the *Relative Consequence Model* proposed by DeLuca et al., (2004), processing speed is proposed to underlie other cognitive deficits, such as working memory. De Luca et al., (2004) propose that working memory is less often impaired in many people with MS, some threshold level of processing speed impairment may be required before it has a functional impact on working memory processes. This might, for example be staved-off by the recruitment of redundant capacity. This was proposed in MS on the basis of a fMRI study of working memory in MS carried out by Chiaravalotti et al (2005), which indicated that even with limited behavioural performance differences with controls, imaging data supported increased activations in additional areas of the brain in achieving those normal levels of working memory performance. Parallel debates as to the key deficit underlying memory decline older people offers some related propositions (Salthouse 1996; Bunce & Macready, 2005), with support for the processing speed decrement than for competing, views such as attentional resource or working memory limitations (Balota, Dolan & Duchek, 2000).

A useful finding from this analysis is the apparent distinction of information processing speed from both working memory and executive function at the factoral level. Interfactor correlations were only moderate (all r's about 0.60), supporting the potential for considering these cognitive functions as to some extent distinct in the sample. Additionally, working memory appears well distinguished from executive abilities with the measures used, with the interfactor correlation of $r = 0.32$ ($p < 0.05$). The final model is shown in Figure 6.8

On balance therefore, the model that best suited the objectives of this study was the four-factor model, pursuing the SDMT as the best indicator of information processing abilities and supported in the MS literature as a sensitive index of this common deficit (Drake et al., 2010). This will therefore be a single indicator, with measurement error modeled, as it is here, for the structural models in the following chapter.

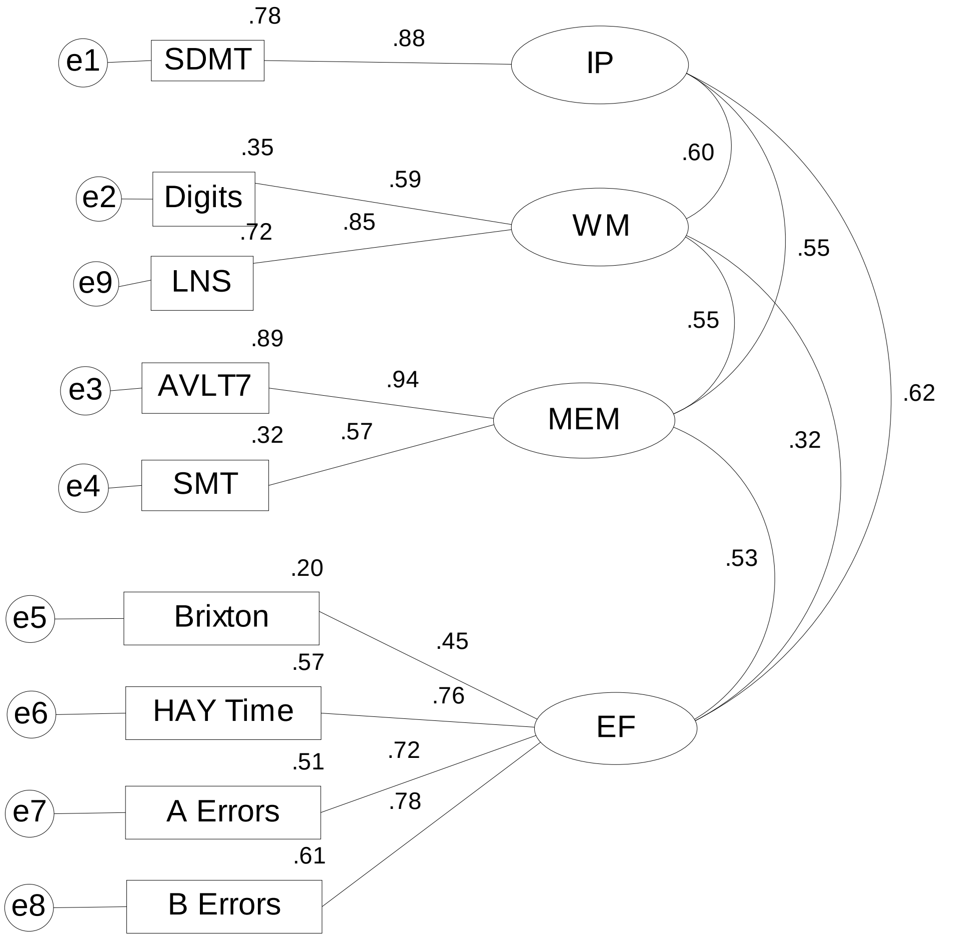


Figure 6.8. The 4-Factor CFA model of Cognitive Items. IP=Information Processing, MEM= Memory, EF= Executive Function, & WM = Working Memory. SDMT=Symbol Digit Modalities Test, Digits=Digit Span Task, LNS=Letter Number Sequencing, AVLT7 = Auditory Verbal Learning Test (trial 7), SMT = Sentence Memory Test, A Errors/ B Errors - A type/ B type Errors Hayling Sentence Completion test. Model fit statistics were: x^2 = 24.370, df = 22 p = 0.328, x^2/df = 1.108. GFI = 0.945, CFI = 0.955 RMSEA = 0.033. Interfactor correlations are all significant at $p < 0.05$.

Table 6.10. 3 Cognitive CFA Models; Memory, Information Processing & Executive Function factors: Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. NPD = Non=positive definite matrix was generated, suggesting poorly specified model. All interfactor correlations are $p < 0.05$. Shaded results are those which present both theoretical *and* statistical fit, representing the accepted four-factor model; *Memory, Executive Function, Working Memory* and *Information Processing.*

Tested Models	**x^2**	**df**	**p**	**x^2/df**	**GFI**	**CFI**	**RMSEA**	**90% CI RMSEA**	**RMSEA p value**	**SRMR**	**Factor correlation**
a priori Model 1	33.772	24	0.089	1.407	0.924	0.816	0.064	0.00 - 0.11	0.30	0.0899	0.63, 0.50, 0.54
Model II								NPD			
Model III	24.370	22	0.328	1.108	0.945	0.955	0.033	0.00 - 0.09	0.617	0.0697	0.53, 0.55, 0.62, 0.60, 0.55, 0.32

6.4. Measurement Models of the Memory Function Questionnaire scales (MFQ)

6.4.1. Forgetting While Reading

6.4.1.1. Introduction

The next measurement model to be tested was that relating to the latent structure of the self-report metamemory measure – the Memory Function Questionnaire (MFQ). The aim of using this questionnaire was to capture self-reported memory performance, or memory self-efficacy (Gilewski & Anthony-Bergstone, 1990; Gilewski, Zelinski & Schaie, 1990; Zelinski & Gilewski, 2004) One issue with carrying out a CFA on this scale was the number of items within the scale, 64 in total. This would generate over 1000 degrees of freedom, depending on the number of factors in the model; not appropriate for a sample of 100 participants.

The scale of interest, proposed to reflect memory self-efficacy is the General Frequency of Forgetting subscale (GFF). This scale contains 33 items and only a portion of this, the section called Forgetting While Reading (FWR) was used as the indicator of metamemory in the Randolph, Arnett & Freske (2004) study. Since that study was published, a 10-item Rasch-modelled version of the 33-item GFF subscale has also been published (Zelinski & Gilewski, 2004). Both scales have 10 items and are therefore considered appropriate for a CFA with data from 100 participants. Factoral structure of each will be tested here. Figure 6.9 summarises how the two scales were derived from the MFQ.

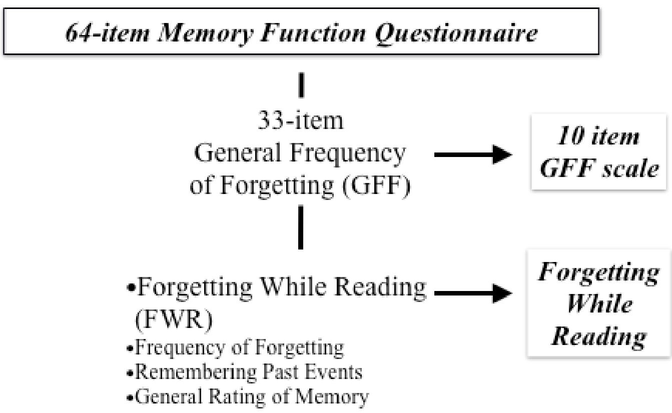


Figure 6.9 Summary of the derivation of the two memory efficacy scales from the Memory Function Questionnaire.

6.4.1.2. Missing Data

11 participants did not fully complete the 64-item Memory Function Questionnaire questionnaire. Checking with participants at the testing appointment that they had not missed out any questions reduced missingness. All other non-responses were considered to be either MAR or non-ignorable. As discussed in the previous chapter, the reasons for not answering related to the question not being relevant to lifestyle or current abilities.

For the Forgetting While Reading scale there was item missingness of 3 out of 100 responses for each of five observed variables. No approach to non-ignorable missingness imputation is without bias, but as there is an expectation of a high level of correlation between all 10 items on this scale and as there is only a very limited amount of data missing, it was decided that a single imputation regression method would be acceptable, instead of the more complex approaches based on ML and MI strategies. The risks of reducing variances and thereby increasing item intercorrelations are considered low (Roth, 1994). In a set of questions such as these, the risks of inflating correlations in already highly correlated items are considered low.

6.4.1.3. Data Screening

The imputed values were examined to confirm they did fall within the expected limits. Two of the 10 items had critical ratios for skew of an absolute value greater than 1.96, and were negatively skewed. Table 6.11 summarises the assessment of normality for the Forgetting While reading items. None of the kurtosis critical ratios exceeded 1.96. In general the direction of skew was as might be expected; more people would report greater difficulty remembering the beginning of a book than the more recent portions (that is give lower scores on memory for earlier portions of the book). The same ‘remoteness’ effect was noted for reports of forgetting while reading a newspaper or magazine. With the limited skew, lack of kurtosis and lack of outlying scores, the data was considered univariate normal.

Table 6.11: Univariate and Multivariate distributions of 10 Forgetting While Reading items from the Memory Function Questionnaire (Zelinski, Gilewski and Anthony-Bergstone, 1990; Gilewski, Zelinski and Schaie, 1990). c.r. = Critical Ratios for Skew and Kurtosis. These are equivalent to Skew and Kurtosis values divided by their standard error, and represents z-scores (Field, 2005: 72). Assuming normality, Skew/Kurtosis have a mean = 0 and a standard error = 1. Critical ratios of > 2.58, plus or minus, indicate a non-normal distribution. Multivariate normality is indicated by Mardia's (Mardia, 1970) coefficient, which also indicates multivariate non-normality when values > 2.58

Variable / Univariate Normality	Skew	c.r.	Kurtosis	c.r.
Forgetting while Reading Novel opening chapters	.169	.690	-.587	-1.199
Forgetting while Reading Novel 3 or 4 chapters earlier	.131	.536	-.698	-1.425
Forgetting while Reading Novel chapter before	-.019	-.078	-.869	-1.733
Forgetting while Reading Novel paragraph before	-.292	-1.193	-.917	-1.873
Forgetting while Reading Novel sentence before	-.513	-2.095	-.875	-1.787
Forgetting while Reading Mag/paper opening paragraphs	-.064	-.263	-.948	-1.935
Forgetting while Reading Mag/Paper 3 or 4 paragraphs	-.093	-.381	-.909	-1.856
Forgetting while Reading Mag/Paper paragraph before	-.225	-.917	-.916	-1.870
Forgetting while Reading Mag/paper 2 or 4 sentences before	-.451	-1.843	-.814	-1.661
Forgetting while Reading Mag/Paper sentence before	-.643	2.627	-.698	-1.425
Multivariate Normality			*38.949*	*12.571*

The critical ratio for Mardia's coefficient was 12.571, indicating a departure from MVN. 15 participants' observations were considered as outlying according to Mahalanobis distance estimates (d^2). However, as individual questions did not demonstrate non-normality, a transformation was not applied.

6.4.1.4.. Model Specification

For the FWR scale a single factor model was proposed, given the potential for collinearity between items with similar questions structure. The 7-choice Likert scales were treated as continuous; the assumption of an underlying continuous distribution is common with categorical responses, especially where there are more than four response options (Bentler & Chou, 1987; Quintana & Maxwell, 1999). There are 10 items in this model, relating to how often memory problems occur while reading a book or newspaper/magazine.

6.4.1.5. Model Testing

The single factor model reflected an assumption that all items reflected a shared dimension. All regression weights were significant and positive as expected, and standardised values ranged from 0.881 - 0.981 suggesting high levels of redundancy, or convergent validity, within the 10-item scale; r^2 values ranged from 0.773 to 0.962. Standardised residual covariances were acceptable, generally < 2.5, suggesting a good fit between model and data-implied covariance matrices. Fit statistics however implied a generally poor fit x^2 = 113.157, df = 35 p = 0.000, x^2/df = 3.233 GFI = 0.771, CFI = 0.330, SRMR = 0.1576, RMSEA = 0.150 (90% CI 0.120 - 0.182; pclose = 0.000).

6.4.1.6. Model Modification

Notable from the median scores is the difference between judgements about earlier parts of book and magazine recall compared to more recent recall. This may be of interest in CFA modelling as it might indicate a shared variance, or redundancy in the questions. Allowing these elements' residuals to covary might therefore be appropriate in order to reflect the application of a *more recent, therefore remember better* heuristic being used to calibrate responses. That is, it reflects some shared variance outside of the measurement in the question itself - which was about *how often* does forgetting take place. In this sense it is proposed to be acceptable to allow covariance between the residuals.

Two residual covariances were applied - between *Magazine/Newspaper, two to four sentences before* & *Magazine/Newspaper sentence before* and between *Novel paragraph before* & *Novel sentence before.* The model was then re-estimated. Again, all regression weights were significant and positive as expected. Standardised values ranged from 0.861 to 0.963. r^2 values ranged from 0.74 to 0.91 and the matrix of Standardised Residual Covariances was acceptable. Fit statistics again implied a generally poor fit x^2= 82.195, df = 33 p = 0.016, x^2/df = 2.491 GFI = 0.834, CFI = 0.578, SRMR = 0.1383, RMSEA = 0.123 (90% CI 0.09 - 0.156; pclose = 0.000). The freed residual correlation paths were significant at r = 0.619 and r = 0.571. Table 6.12 summarises the initial and modified model fit statistics.

A two-factor solution was also investigated, based on a 'Book/Novel' items and 'Newspaper/Magazine' items, but with an inter-factor correlation of 0.89 as evidence of unidimensionality, it was decided that a single factor better reflected convergent validity and that a two factor solution offered no gains in discriminant validity.

The Forgetting While Reading CFA presented generally poor fit statistics, which is likely to relate to the impact of item-word redundancy, a common-methods effect (Campbell-Sills & Brown, 2005). This is where similarity of questions leads to higher covariances between items. While wanting correlations between items indicated by the same underlying factor, sometimes issues with the fit of the CFA such as experienced here can arise because of very high correlations (Netemeyer, Bearden & Sharma 2003). Items that are highly correlated and share variance may also present correlated error (residual) terms (as tested in the modification stage above). The fact that residuals were correlated, points to common-methods effects as one probable source of poor model fit and supports a redundancy rather than convergent validity interpretation (Netemeyer, Bearden & Sharma 2003).

6.4.1.7. Summary

It is recommended that CFA models be based, at least to some extent, on previous exploratory analysis (Floyd & Widaman, 1995) and the Forgetting While Reading scale was not derived as a single factor in the proposed four-factor structure of the original Memory Function Questionnaire. Instead, it was part of a 33-item General Frequency of Forgetting scale, proposed to reflect memory self-efficacy. This scale was recently reduced further to a 10-item scale (the GFF-10) by rasch modelling (Zelinski & Gilewski, 2004) and it is this that will be submitted next for CFA. It is proposed that this latter scale has a psychometric provenance that makes it a more appropriate latent variable to be used to measure memory self-efficacy.

Table 6.12. Forgetting While Reading Models; Two assessments for single factor Forgetting While Reading model. The initial model represents a single factor, the modified model a single factor model with error correlations to reflect the similar question structure of questions about reading books and magazines. Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual.. All residual correlations are $p < 0.05$.

1 Factor FWR models	x^2	**df**	**p**	x^2**/df**	**GFI**	**CFI**	**RMSEA**	**90% CI RMSEA**	**RMSEA p value**	**SRMR**	**Residual correlation**
Model 1 - single FWR factor	113.157	35	0.00	3.233	0.771	0.330	0.150	0.012 - 0.182	0.00	0.1576	n/a
Modified Model - single FWR factor with correlated residuals	82.195	33	0.016	2.491	0.834	0.578	0.123	0.09 - 0.156	0.00	0.1383	0.619, 0.571

6.4.2. 10-item General Frequency of Forgetting Measurement Model

Alternative CFA models of the GFF-10 were tested also, because of the original, and diverse, 33-item scale from which the 10 items were drawn is made up of a number of subscales. This suggests some potential for factoral separation might be possible. Tested models comprised a single-factor model and a three-factor model.

6.4.2.1. Missing Data

In the second dataset (GFF-10), there were again 10 observed variables with 100 observations of each. For these, there was an item missingness of 3% for 2 questions. A single regression-based imputation was therefore applied here also.

Normality.

All values for skew and kurtosis were considered acceptable with absolute values < 2.0. The critical ratio for Mardia's coefficient was 4.390, indicating some departure from MVN. Assessment of multivariate outliers using Mahalanobis distance suggested approximately 5 outlying respondents' scores. Generally therefore, there was some departure from multivariate, though not univariate, normality in the response distribution.

6.4.2.2. Model Specification

1-Factor Model of GFF-10

Based on the findings of Zelinski & Gilewski (2004), a single factor model of the 10 items was first tested.

6.4.2.3. Model Testing

Regression weights were all positive and significant at p < 0.001, standardized estimates ranged from 0.479 to 0.961 with Standardized Residual Covariances indicating some discrepancies between model-implied and data-implied model fit; r^2 values ranges from 0.23 (*Faces*) to 0.92 (*Novel Reading: forgetting last paragraph*). The fit statistics for the model were x^2 = 93.336, df = 35 p = 0.00, x^2/df = 2.667 GFI = 0.811, CFI = 0.379, SRMR = 0.1715, RMSEA = 0.130 (90% CI 0.098 - 0.162; pclose = 0.00). General fit statistics did not therefore support this model as a good fit for the data.

6.4.2.4. Model Modification

Modification indices suggested an error covariance for *Remembering Faces* and *Remembering Names* so the model was re-estimated after inclusion of this, with the following values obtained; x^2 = 80.081, df = 34 p = 0.00, x^2/df = 2.377. GFI = 0.837, CFI = 0.502, SRMR = 0.1618, RMSEA = 0.118 (90% CI 0.058 - 0.151; pclose = 0.0) Standardised residual matrix values remained unacceptable for some items, but regression weights were both significant and acceptable.

As before, because of the potential for common methods effects, residual terms for the two *Remembering While Reading* items, and two items asking '*How well do you remember things that occurred…' 1-5years ago* or *between 6 and 10 years ago*, were also allowed to correlate. This gave a final model with 3 sets of shared error variances and fit statistics of; x^2 = 50.744, df = 32 p = 0.019, x^2/df = 1.586 GFI = 0.897, CFI = 0.800, SRMR = 0.1107, RMSEA = 0.077 (90% CI 0.032 - 0.115; pclose = 0.136). The matrix of standardised residuals was acceptable and all parameters were significant and positive in value. Considerable benefits to model fit were gained from the addition of error (residual) covariances, though none of the versions of the single factor were satisfactory across the full range of fit indices. Appendix J presents the full data from modelling, and the initial and 2 modified versions of the 1-factor GFF model fit assessment is presented in Table 6.13

An alternative interpretation is that the three residual term covariances actually indicate three separable factors, and since residual covariances would not be represented in the factor scores used in structural modelling, a 3-factor model of this scale was constructed.

3 Factor Model of GFF-10

This model consisted of a factor structure reflecting the original derivation of the scale; Forgetting While Reading (FWR), Retrospective Functioning (RF) and Frequency of Forgetting (FF). Results were: x^2 = 59.906, df = 32 p = 0.002, x^2/df = 1.872 GFI = 0.879, CFI = 0.703, SRMR = 0.1252, RMSEA = 0.094 (90% CI 0.056 - 0.130; pclose = 0.032). Regression weights were all significant and positive and the matrix of standardised residuals was acceptable. Interfactor correlations were r = 0.675, 0.621 and 0.569. Although the x^2 was non-significant, indicating poor fit, the 'normed x^2' (x^2/df ratio) was < 2.0, suggested by some to be a good index of fit in models where x^2 might be limited by sample size or departures from multivariate normality (Carmines & McIver, 1981; Byrne, 2001; Kline, 2005). Summary fit statistics are presented in table 6.14, the full AMOS outputs in Appendix K and the model is shown in Figure 6.10

Table 6.13. 10-item General Frequency of Forgetting (GFF-10) Scale Models; Two assessments for single factor Forgetting While Reading model. The initial model represents a single factor, the modified model a single factor model with error correlations to reflect the similar question structure of questions about reading books and magazines. Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. All residual correlations are $p < 0.05$.

GFF-10 1-Factor Models	x^2	df	p	x^2/df	GFI	CFI	RMSEA	90% CI RMSEA	RMSEA p value	SRMR	Residual correlation
Initial Model	93.336	35	0.000	2.667	0.811	0.379	0.130	0.098 - 0.162	0.000	0.1715	n/a
Modified Model	80.081	34	0.000	2.377	0.837	0.502	0.118	0.058 - 0.151	0.000	0.1618	0.44
Final Model	50.744	32	0.019	1.586	0.897	0.800	0.077	0.032 - 0.115	0.136	0.1107	0.47, 0.81, 0.59

Table 6.14. 3-Factor GFF: Summary of goodness of fit statistics. x^2=chi square value; df=degrees of freedom; x^2/df = chi square to degrees of freedom ratio; GFI = goodness of fit index; CFI = comparative fit index; RMSEA= Root Mean Square Error of Approximation; 90% CI RMSEA= 90% confidence intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. All interfactor correlations are $p < 0.05$.

3 Factor GFF Model	x^2	df	p	x^2/df	GFI	CFI	RMSEA	90% CI RMSEA	RMSEA p value	SRMR	Factor correlation
Initial Model	59.906	32	0.002	1.872	0.879	0.703	0.094	0.056 - 0.130	0.032	0.1252	0.68, 0.62, 0.57

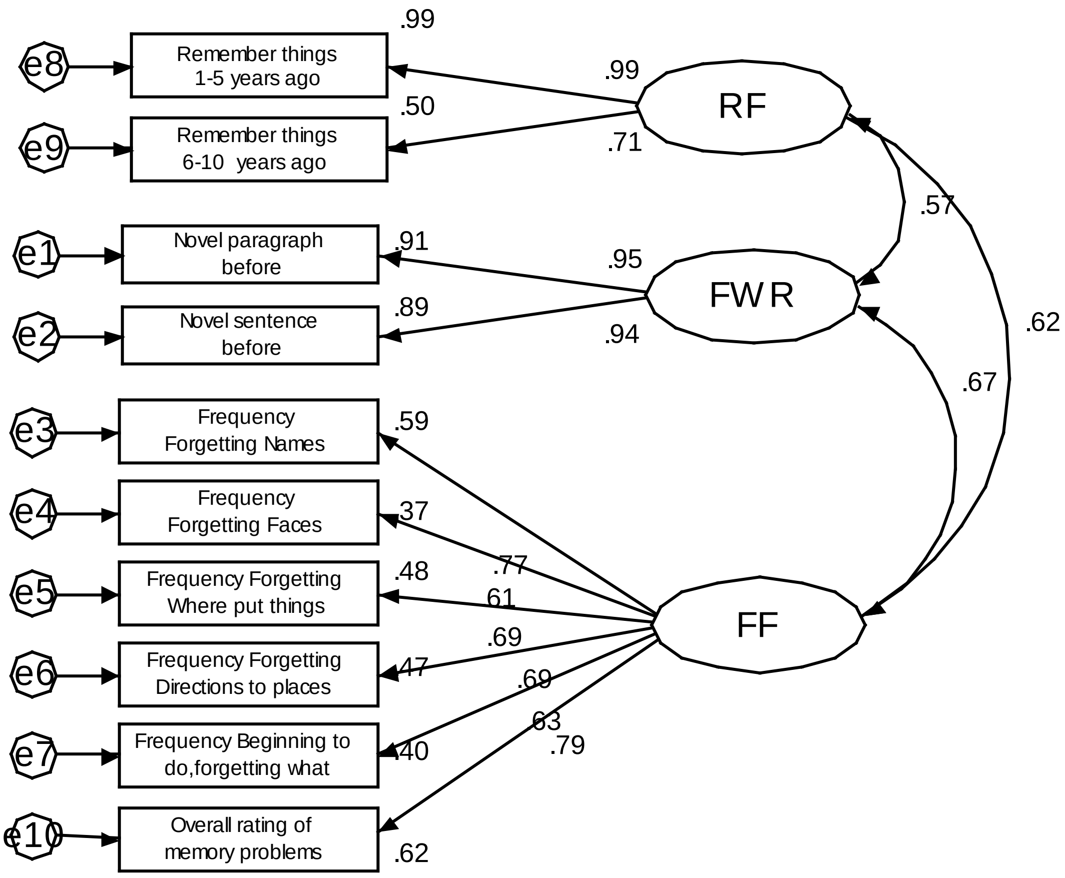


Figure 6.10 3 Factor General Frequency of Forgetting Model; Factor 1 = Retrospective Functioning. Factor 2 = Forgetting while Reading and Factor 3 = Frequency of Forgetting. Fit statistics were x^2 = 59.906, df = 32 p = 0.002, x^2/df = 1.872 GFI = 0.879, CFI = 0.703, SRMR = 0.1252, RMSEA = 0.094. Each factor correlated $p < 0.05$

6.4.2.6. Discussion.

The three-factor GFF model lends some support to the proposed structure of the larger questionnaire from which it was drawn, where each factor was a separate scale - *Retrospective Functioning, Forgetting While Reading & Frequency of Forgetting* (Gilewski, Zelinski & Schaie, 1990). It also offers gains in terms of absolute fit indices.

There are limitations in all of the MFQ derived measurement models, and in larger samples they could be investigated for pruning of congeneric items. While the tools internal consistency was reported to be high (> 0.80) for all factors (Gilewski, Zelinski & Schaie, 1990), this analysis suggests that this may relate to item redundancy, especially for *Forgetting while Reading* items. Alternatively, unlike the cognitive and even mood latent variables derived by CFA, the concept of memory self-efficacy is unclear, and this too may be reflected in the results of this factor analysis.

In accepting a single factor for the *Forgetting while Reading* and a three-factor model for *General Frequency of Forgetting* these limitations are acknowledged. For structural modelling, in part to reduce the numbers of parameters to be estimated, the factors will be reduced to factor scores.

In respect of the scales themselves, the lower incidence of missingness in the GFF, compared to the FWR, scale may suggest more validity as a measure of self-efficacy for people with MS, for whom reading may not be a reliable way to index subjective memory impairment, being that it may be heavily confounded with other aspects of the disability, such as vision and upper limb function.

6.5. General Discussion of Measurement Models

Confirmatory factor analyses have been reported here to test the measurement models for cognitive, mood and memory-efficacy measures. The aim in testing the measurement model was to clarify the contributions of observed variables to latent variables, for the next stage of analysis. The model of cognitive items supported a four, rather than three, factor solution, underlining differences between information processing speed and working memory, and in disagreement with the proposed relationships of the *a prior* model.

However, this is in line with some of the literature in MS that suggests information processing is a generally imprecise latent construct. Because of the orientation of the study, information processing will be indicated in structural modelling by the Symbol Digit Modalities Test, and its modelled measurement error. Working Memory will be dropped from further analysis. This maintains the focus of the study on the mediating role of information processing on effortful metamnemonic judgments.

In respect of measurement models for both mood and memory-efficacy, both perhaps suffered from the combined effects of restricted measurement scale, some non-normality in their distribution, similarity in questions leading to high levels of collinearity in the CFA models, which appeared to compromise fit. The BDI, using transformed data, presented a higher than ideal number of to-be-estimated parameters for the sample size, though an acceptable fit was found, in line with a 2-factor model proposed by the test authors (Beck, Steer & Brown, 1996). This CFA is the first analysis of this tool in a sample of people with MS, and has utility for examining the relative impact of the disease on somatic aspects of depression, as compared to the more cognitive and affective components. In terms of factor loadings, *Worthlessness* is the larger factor loading for the Cognitive/Affective factor and *Loss of Concentration* and *Lack of Energy*, for the Somatic factor.

To summarise, latent variables have now been confirmed; *Memory* is indicated by delayed recall on the Sentence Memory Test and the delayed recall trial of the Auditory Verbal Learning Test and *Information processing* is indicated by the Symbol Digit Modalities Test. *Executive Function* is indicated from three Hayling measures and one measure from the Brixton test. For the *Mood* latent, two factor scores will be created to reduce the number of parameters in structural models. Similarly, for the two memory-efficacy scales - one factor score for FWR and three for GFF.

Chapter 7: The Structural Models.

7.1. Introduction

This final chapter of results reports the findings of an assessment of a number of structural equation models exploring the relationships between the derived cognitive and mood latent variables, and metamemory measures; Memory self-efficacy, Judgment of Learning, Retrospective Confidence Judgment and Feeling of Knowing. The generalised *a priori* models were outlined in Chapter 1, and the results of their testing are reported here, including extensions to those models, suggested both by findings in the previous results chapters and indicated modifications from initial model fit statistics. In this regard the *a priori* model begins an iterative process. The models specified in Chapter 1 have been refined in three respects. The selection of accuracy measures for metamemory judgments was refined to include both relative and accuracy measures for Retrospective Confidence Judgment and Feeling of Knowing. Secondly, the testing of measurement models, reported in the previous chapter means that the latent variable are indicated differently to the a priori measurement models. Finally, the results from Chapter 5, notably for Judgment of Learning will inform model specification amendments.

For each result, as in the previous chapter, a model is specified, tested and modifications considered based on results. These modifications are used in later models to advance understanding of relationships between latent contributors and metamemory. Later models are more exploratory, based on relationships that might be expected from the proposed underpinnings of metamemory judgments. Informing model specification too is an awareness that in MS, models may offer affirmative or contradictory findings from the metamemory theory base, itself mainly derived from neurologically-intact individuals. A concurrent aim is consideration of the contribution to information processing speed and mood as relevant factors in the more effortful aspects of metamemory judgment.

A direct mapping of the latent construct of *Memory* to all mnemonic cues, and *Executive Function* to all inferences is not proposed; the literature on metamemory is not clear in suggesting this distinction can be cleanly made (Nelson et al., 1984; Koriat et al., 2008; Leonesio, 2008), though a dual process is proposed - reflecting what Koriat (2007) has termed information-based and experience based processes, what others have termed automatic versus effortful (Koriat et al., 2008) and others, trace-access and inferential mechanisms (Nelson et al, 1984; Leonesio, 2008). As discussed in previous chapters, task-based metamemory will be measured from two perspectives of accuracy - absolute and relative. The approach used for absolute accuracy is a focus separately on accuracy and inaccuracy. Model interpretation will explore relationships between cognitive and mood latent variables and inaccuracy, with the assumption that inaccuracy has clinical relevance as a target for intervention. Understanding the factors that contribute to it is therefore of relevance also. In the absence of memory-experience for example, inferential judgments might contribute to inaccuracy (Leonesio, 2008), or affect bias efficacy judgments (Randolph, Arnett & Freske, 2004).

The structure of the chapter is as follows; a specification of each model will be outlined, including the expected parameter sign and significance. The model will be then tested and modification reviewed based on the result. As before, a set of fit indices will be considered to assess model acceptability, in conjunction with parameter size and sign, and its acceptability in providing a theoretically plausible account of the data. Final models will be selected for each judgment and accuracy type, with only initial discussion of each result presented. All models are reported using standardised parameter values.

Where possible, a number of questions will be addressed in a single model assessment. One of the advantages of the SEM approach is that a number of simultaneous comparisons can be made, avoiding the need for multiple regression-based parameter estimations (Iacobucci, 2008) and providing better empirical support for findings because the results can be interpreted in the context of all other relationships in the model. So, as before, *specification*, *testing*, *modification* and *re-testing* is presented, followed by evaluation and initial interpretation. As with the previous chapter, this chapter integrates a discursive approach because the modelling process is as much a theoretical endeavour, as it is statistical. A convention of indicating a model's latent variables will be followed wherein they will be italicised, whereas the domain they are proposed to represent will not; this will mean *Memory* refers to a latent variable used in the model whereas 'memory' relates to the cognitive domain.

7.2. Memory self-efficacy.

7.2.1. Model Specification

There are a number of objectives in this set of results, based on the findings of Randolph, Arnett & Freske (2004), that non-memory cognitive processes might be mediated by negative attitude or affect in efficacy judgments about memory. Beatty & Monson (1991) in respect of MS, and many others more generally (Hertzog, 2002), propose that tested memory ability is generally not associated with these efficacy judgments.

Here, given the availability of some relevant literature, it is appropriate to test all aspects in one model, and assess whether the combined theory of published studies can be supported.

The aims addressed in model testing are to:

1. Validate the model of Randolph, Arnett & Freske (2004) that performance on a metamemory scale is contributed to by both mood and executive function, with mood acting as a mediator of executive relationships with memory self-efficacy.
2. To confirm that subjective memory report is not associated with tested memory performance. This will be tested by the inclusion of *Memory* in the self-efficacy models, with the expectation of finding non-significant parameter estimates.

The key expected relationships in the models are:

1. A negative association between executive ability and mood disorder and between mood disorder and metamemory, because more depression will be associated with lower executive ability and lower questionnaire scoring (indicating more memory complaint).
2. The direct effect of *Executive Function* on Metamemory, in the context of *Mood* mediating *Executive Function*, is expected to be non-significant, so indicating no direct mapping of executive abilities to memory self-efficacy, in line with the findings of Randolph, Arnett & Freske (2004).
3. *Memory* and *Executive Function* are expected to correlate to a significant extent

The validation of the findings of Randolph, Arnett & Freske (2004) will be tested with the two measures of memory self-efficacy previously outlined; the 10-item General Frequency of Forgetting scale, and the Forgetting-While-Reading scale, the latter used by Randolph, Arnett & Freske, the former proposed as a more valid measure of memory self-efficacy (Zelinski & Gilewski, 2004). The model, reflecting these aims, and expected relationships, is presented below in Figure 7.1.

Given that two variables have a large number of observed items (Beck Depression Inventory has 21, FWR and GFF each have 10 items) cognisance had to be taken of the ratio of observed items, and therefore estimated parameters, to the sample size for the study. It is proposed here that the structural models benefit from having the confirmed measurement structure established in Chapter 6, prior to model testing. As a result of the large numbers of indicators for these scales, factor scores are created for those latent variables.

Factor scores are considered potentially useful, and intuitively straightforward, ways of testing the interaction of latent variables (Marsh, Wen & Hau, 2006). However, they suffer from some of the limitations of using observed variables, in that they effectively remain a conflated measure (that is, measurement error and true score combined) in the psychometric context. Thus, as Bollen (1989) discusses, the key limitation, after partialling measurement error from true latent-related variance during CFA, factor scores effectively re-combine both types of variance together to create a score. Some suggest that the use of SEM and CFA negates the need to pursue factor score creation at all (Brown, 2006). Here, it is seen as a pragmatic approach to reducing model complexity (Grover & Vriens, 2006), because the numbers of free parameters would be inappropriate for model testing with a sample size of 100 (Boomsma & Hoogland, 2001). Based on the confirmatory factor analyses presented in the previous chapter, regression-derived factor scores were therefore created for the two Beck Depression Inventory factors (Cognitive/Affective and Somatic) and for the Forgetting While Reading and GFF-10 scales. In the models to follow, there are 2 factors for the Beck Depression Inventory, 1 factor score for the Forgetting While Reading items, and 3 factor scores for the General Frequency of Forgetting 10-item scale.

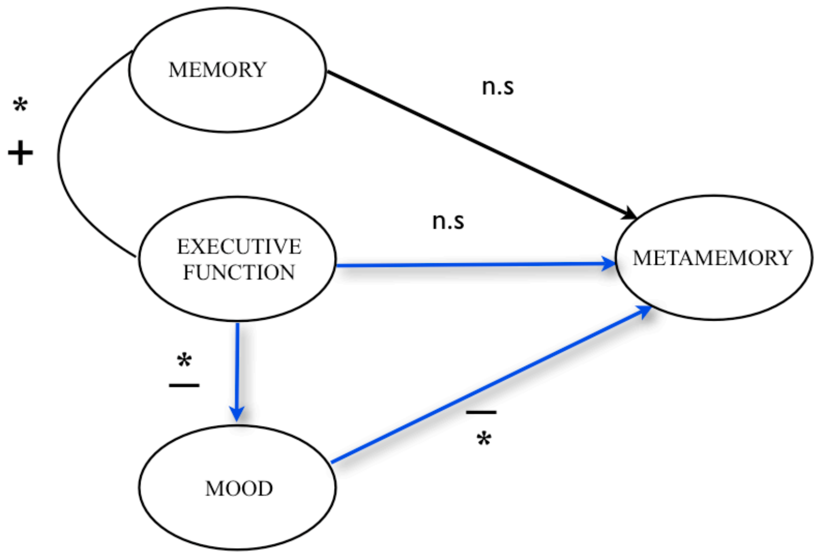


Figure 7.1: Memory Self-efficacy model including expected parameter signs (+ and -) and significant paths; **n.s.** = expected non-significant paths; ***** = expected significant paths. The mediational relationship is outlined in blue. Observed measures contributing to latent variables are not shown.

Setting residual variance.

Because a single factor score was created for the FWR indicator of the metamemory latent variable, a residual value (error variance) was set, based on the reliability of the MFQ as a whole (Gilewski, Zelinski & Schaie, 1990; Zelinski, Gilewski & Anthony-Bergstone, 1990); this was required both in the interests of making an identifiable model, and to reflect measurement error, related to the reliability of the FWR variable. The setting of this residual, or error variance, is based on the formula provided by Bollen (1989) and is given by:

$$(1-r_x)var_x \qquad \textit{formula 7.1}$$

Where r_x is the reliability of the measure used, and var_x the variance within the data derived from the measure. The residual value for FWR was 0.1563.

Assessment of mediation.

A mediational mechanism in which *Mood* mediates the relationship of inferential judgment and self-efficacy is also proposed in this model, reflecting findings in the literature. The process for assessment has been discussed in Chapter 3: *Development of Statistical Methods*. The specific mediational relationship tested here is outlined in figure 7.2. To summarise, the Sobel *z*-test is carried out to test if the difference between path, *c* as a direct effect and *c'* as a direct effect with the addition of the mediator, is statistically different from zero. If it is, some mediation is proposed to exist.

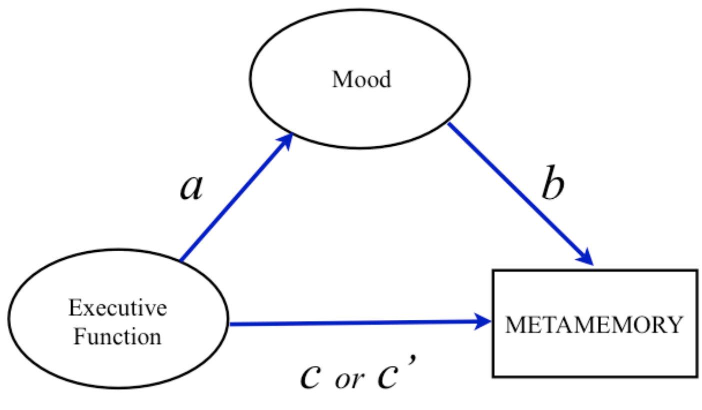


Figure 7.2: A schematic model of mediation of *Executive Function* by *Mood*. The test of mediation assesses whether the difference between path c, the direct effect of Executive Function on Metamemory and path c' the effect when mediation is also modelled, is statistically different from 0, so that a*b should be equal to c-c' if mediation is not extant.

7.2.2. Model Testing Results

Initial model testing generated a negative error variance (Heywood case) for the Somatic factor loading on *Mood*. This latent variable was estimated with the Cognitive/Affective and Somatic factors from the BDI-II measurement model. The standardised regression weight for the loading of the Somatic factor on *Mood* was greater than 1.0 in this model estimation, caused by a negative error variance of -0.059 SE = 0.282, that is the variance is not statistically different from zero. It can occur because of multicollinearity between indicators or almost perfect linear relationships between the observed variable and its latent. Error variance for this item was set to a non-zero value (0.01) in line with recommendations of Joreskog & Sorbom, (1999). The Heywood case may reflect the creation of this variable from factor scores, which include measurement error.

The GFF-10 model, had the following global fit results: x^2 = 49.035, df = 40, p = 0.155, x^2/df = 1.226, GFI = 0.910, CFI = 0.876, RMSEA = 0.048 (90% CI = 0.00 - 0.088; pclose = 0.503), SRMR = 0.0961. *Memory* and *Executive Function* correlated as expected and to a significant extent (r = 0.65, p = 0.003). The model was acceptable in terms of fit statistics, parameter sign and size. Only *Mood* had a direct and significant effect on efficacy (*B* = -0.408, p = 0.008), indicating that participants who reported higher depression also reported more memory difficulties (lower efficacy scores). This was in the absence of a relationship between memory function and *Metamemory*. A Sobel test of mediation was *z* = 1.598, suggesting that the relationship between executive function and *Metamemory* was mediated, by *Mood*. 23% of the variance of metamemory was explained in the model. Appendix L presents the full statistical outputs related to this model. Figure 7.3 Summarises The Model And Findings.

Ideal model fit statistics are summarised in Table 7.1, and is used for the assessment of global fit of each model tested. Additional assessments of model fit are outlined in Table 7.2.

Table 7.1: Fit indices and other assessments of model acceptability. x^2 = Chi Square value; x^2 p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual. Interfactor correlations are relevant to the interpretation of Confirmatory Factor Analyses with values > 0.80 suggesting that factors may be too closely relate to be discriminated.

Specific indices of model fit	Other Assessments
x^2 & p-value	Standardized Residual Matrix values generally < 2.58
CMIN/df	Interfactor correlations in 2+ factor models < 0.80
GFI	Squared Multiple correlation *(r^2)* values > 0.20
CFI	Parameter sign & significance
RMSEA: 90% CI & p-value	
SRMR	

Table 7.2: Summary of recommended values for model fit indices. x^2 = Chi Square value; x^2p-value = significance of chi-square statistic; CMIN/df = chi-square to degrees of freedom ratio; GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA=Root Mean Square Error of Approximation; 90% CI = 90% confidence Intervals for RMSEA; SRMR = Standardised Root Mean Square Residual

x^2 p-value	CMIN/df	GFI	CFI	RMSEA	SRMR
p > 0.05	< 2.0	> 0.90	> 0.90	90% CI upper limit of 0.08; pclose > 0.50	< 0.10

The FWR model (figure 7.4) generated similar conceptual results; mediation of executive function by mood support for the relationship between mood and metamemory, and no support for a relationship between tested memory and memory efficacy. Global fit was $x^2 = 36.125$, df = 24, p = 0.053, x^2/df = 1.505, GFI = 0.919, CFI = 0.813, RMSEA = 0.071 (90%CI = 0.00 - 0.117; pclose = 0.219), SRMR = 0.0944. 16% of the variance in metamemory was explained by the model. A Sobel test of mediation was $z = 1.748$ and using the MacKinnon, Lockwood & Williams, (2004) cut-off ponts, indicating mediated relationship. Appendix M presents the full statistical outputs related to this model

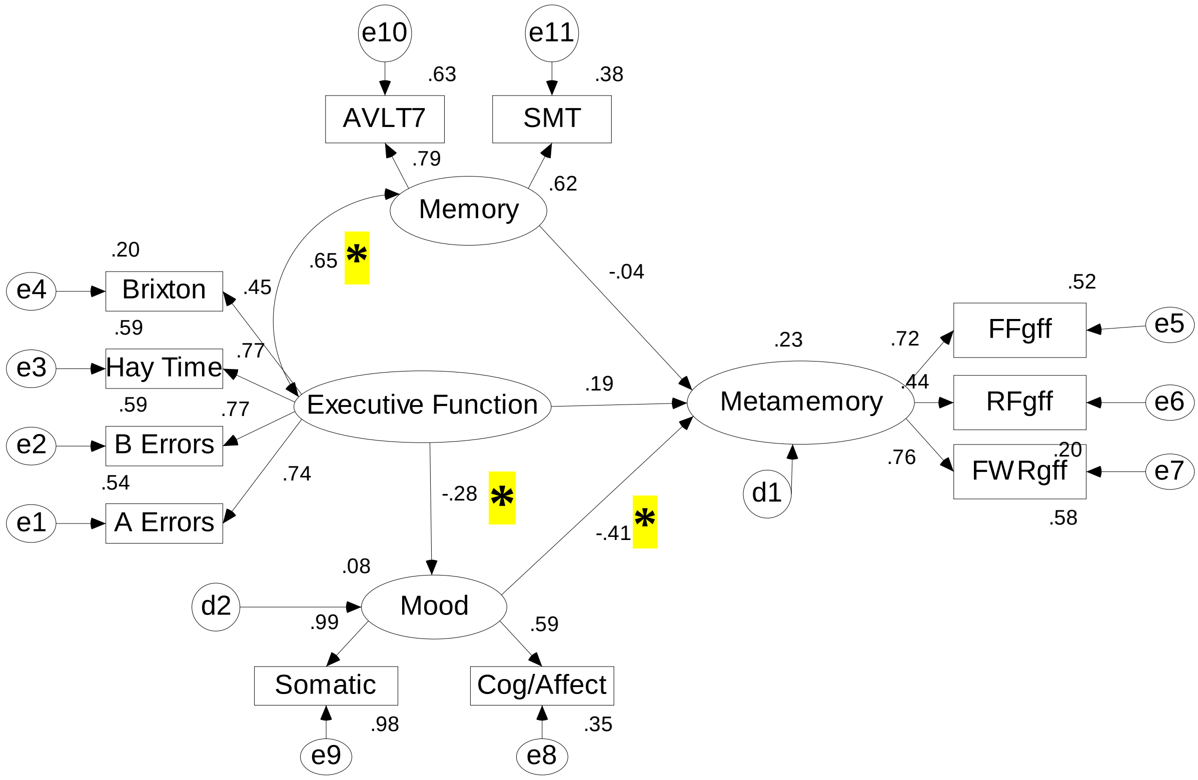


Figure 7.3: The mediation of *Executive Function* by *Mood* in contributing to *Metamemory - General Frequency Of Forgetting*. Metamemory is described by the General Frequency of Forgetting scale factors * = significant path at $p < 0.05$ level. *Memory* was not a contributor in this model. All latent variable loadings were also significant. FFgff = Frequency of Forgetting items (factor score) from the General Frequency of Forgetting scale; RFgff = Retrospective Functioning items (factor score) from the General Frequency of Forgetting scale; FWRgff = Forgetting while Reading items (factor score) from the General Frequency of Forgetting scale; AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items, and d = residuals (disturbance) for endogenous latent variables.

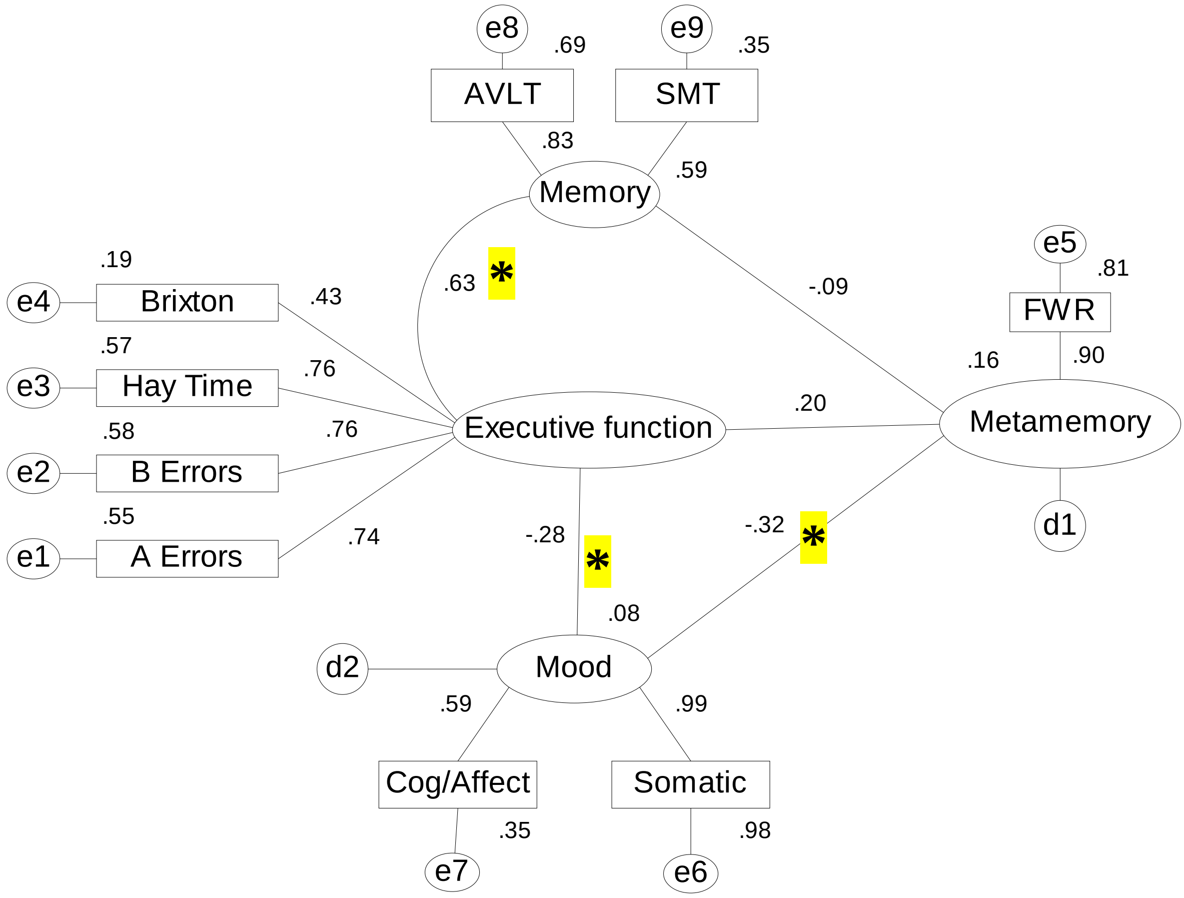


Figure 7.4: The mediation of *Executive Function* by *Mood* in contributing to *Metamemory*. Metamemory is described by the Forgetting While Reading scale factor score. * = significant path at $p < 0.05$ level. *Memory* was not a contributor in this model. All latent variable loadings were also significant. FWR = Forgetting while Reading factor score; AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score).e = residuals for observed items, and d = residuals (disturbance) for endogenous latent variables. measurement error was modelled for the single item Metamemory latent variable.

7.2.3. Summary of memory self-efficacy models.

Model fit indices were generally as desired, with only the Comparative Fit Index (CFI) extending below the ideal values, notable in the Forgetting While Reading (FWR) model. The CFI generally reflects the average level of correlations within the models, is related to degrees of freedom, so that each additional parameter estimated (or variable added) is penalised (Schumacker & Lomax, 2004). In part, the inclusion of *Memory* is likely to have reduced its value therefore, without contributing to the model.

Parameter estimates, in terms of sign and significance were also as expected, so supporting the *a priori* models and confirming the findings of Randolph, Arnett & Freske (2004). Mood was supported as the only direct contributor to *Metamemory* in both models, though fit statistics were more acceptable using the more valid instrument, the General Frequency of Forgetting scale. *Memory* did not relate to subjective memory appraisal, in agreement with the findings of Lovera et al., (2006) and Julian, Merluzzi & Mohr (2007) that self-assessed cognitive dysfunction in MS typically relates more to depression, than actual cognitive abilities. Mood may also mediate the more inferential judgment involved in this type of metamemory assessment.

7.3. Task-based Metamemory Judgments

7.3.1 Introduction

The aims of model testing here are to address the recommendations for further study laid out by Pannu & Kaszniak (2005) to better understand metamemory in neurological populations by assessing a number of different measures of metamemory. The order of analysis is structured around the stages of memory at which each judgment is made; Judgment of Learning (JoL) at acquisition, Retrospective Confidence Judgment (RCJ) at recall, and Feeling of Knowing at recognition. The Feeling of Knowing model will be assessed against the Beatty & Monson's 1991 data on the performance of people with MS.

The goals of modelling are twofold; first, to examine direct contributions of memory, executive function and affect to metamemory. This will be in the form of a direct effects model for each judgment. The informing theory has been discussed, but generally would suggest that mood is unrelated to task-based judgments. Studies not finding such a relationship however are in samples where memory, executive function or mood is not compromised. It is possible mood may inform judgments where there are difficulties in cognitive aspects of the task, or where efficacy-related inferences are used in the presence of cognitive deficit. However, the relationship between mood-disorder and metamemory in MS has not been investigated for this judgment. For this reason, *Mood* will be retained for assessing any direct contribution to metamemory judgments.

Second, mediation is investigated with information processing as mediator of proposed effortful processes in making metamemory judgments. Information processing is suggested as a mediator to many cognitively effortful tasks; some of the judgments modelled here can be considered effortful, notably those judgments that are not typically considered to be based alone on retrieval fluency; one such is the Feeling of Knowing judgment. Both Judgments of Learning and Retrospective Confidence are likely more supported by memory experience than Feelings of Knowing because they involve judgments made on items for which some retrieval has already taken place. Feeling of Knowing judgments are based on items for which there has been a retrieval failure, perhaps making them less dependent on processing related to explicit memory experience. To this end, it is proposed that of primary interest are results for *Executive Function* being mediated by information processing in Feeling of Knowing models. For each however, the contribution of executive-level processing is considered to be the more controlled type and therefore likely to be mediated by information processing. These contentions will be tested and expectations for each model will be outlined prior to testing. The overall aim of the next series of models then, is to assess mnemonic and executive contributions to metamemory accuracy, to assess whether information-processing abilities facilitate these contributions and to enquire is mood disorder relates to accuracy in task-based metamemory judgment.

7.3.2. Judgment of Learning.

Judgments of Learning at delay continue to provide debate both in terms of their utility, and the basis on which they are made (Meeter & Nelson, 2003; Kimball & Metcalfe, 2003; Dunlosky& Bjork, 2008; Narens, Nelson & Scheck, 2008; Dunlosky & Metcalfe, 2009). One important component of this debate is whether they reflect memory or metamemory. Additionally, whether they are based on direct-access, mnemonic experience, or inference made from normative aspects of the task such as task difficulty is debated (Dunlosky & Bjork, 2008). It was proposed in developing the forthcoming model that both *Memory* and *Executive Function* would encompass both types of contribution.

7.3.2.1. Model Specification

Delayed Judgment of Learning (JoL) was based on the 20-minute delayed recall trial of the Auditory Verbal Learning Test, and was calculated as the discrepancy between predicted and actual recall performance. Accuracy in the judgment is indicated by the size of the discrepancy between the two. A score of zero indicates a well-calibrated judgment and departures from that, increasing inaccuracy. The *meta* contribution is therefore contained in the discrepancy, not the memory performance. Predicted score was lower than actual score for 95% of participants for the *delayed* JoL, so values were typically negative, indicating underconfidence.

Overall performance of the sample has been outlined in Chapter 5, and is summarised in Figure 7.5, which indicated that for the sample, there was evidence of an *underconfidence with practice effect* (Koriat, Sheffer & Ma'ayan, 2002). This effect is characterised by an initial small overconfidence, followed by increasing underconfidence with repeated learning trials (Meeter & Nelson, 2003). The effect has been demonstrated in both item-by-item judgments and global judgments (Dunlosky & Metcalfe, 2009; Koriat, Sheffer & Ma'ayan, 2002). In many studies, at the *delayed* JoL stage, the underconfidence is extinguished, so that the accuracy of the *delayed* JoL is greater, compared to prior trials' JoL. Even in populations with neurological impairment, this delayed Judgment of Learning effect has been noted (Pannu & Kaszniak, 2005). As reported in Chapter 5, the mean JoL was less accurate after with delay in this sample, compared to the last of a set of immediate JoLs (Trial 5); $t\ (99) = 2.148$, $p = 0.031$. It is of particular interest here that the delayed Judgment of Learning effect was not found, a result that has implications for model specification.

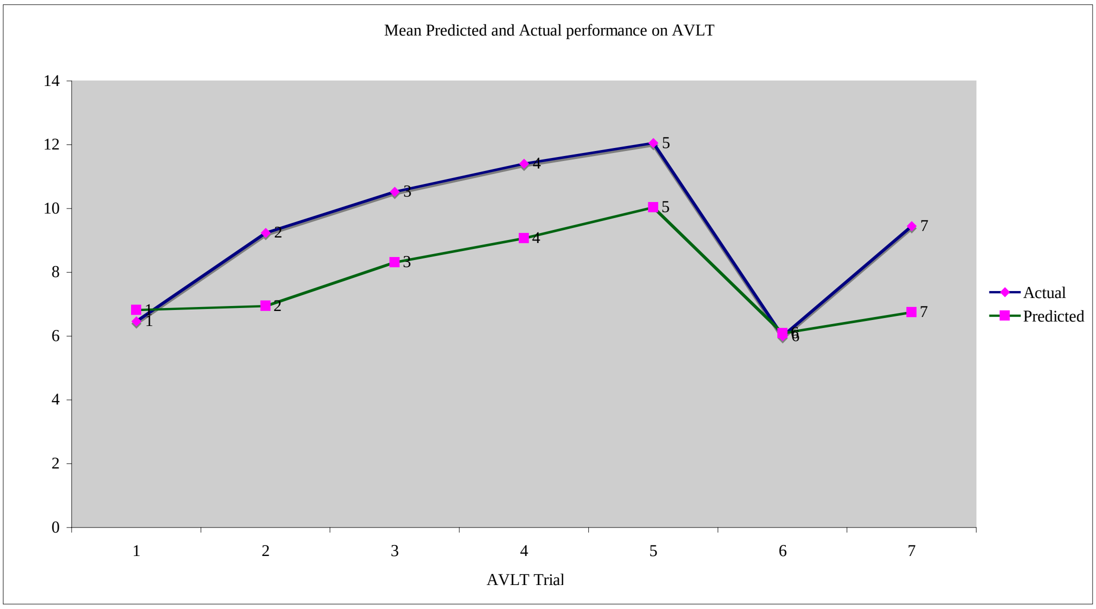


Figure 7.5 Summary of Judgment of Learning performance across trials of the Auditory Verbal Learning Test. Serial prediction is presented in green with actual performance in blue, both indicating mean prediction and performance on each tests of the 15-word list. Trial 6 is a new list episode and Trail 7 a 20 minute delayed recall test.

In one study on neurologically unimpaired participants, Meeter & Nelson (2003) report a similar finding to this study, and suggest that the failure to ameliorate underconfidence with delay in making the judgment might relate to insufficient weight being given to external cues, such as the number of repetitions in the task, or incorrectly focusing on the wrong cues to make the decision. One such cue could relate to the use of the pre-delay JoL to guide the subsequent delayed JoL (Finn & Metcalfe, 2007). This finding was therefore considered in specifying the model. The proposal that previous trial judgment contributes to later judgment, was not considered initially, but was added because the continued underconfidence after delay was a novel finding.

Three features of Judgments of Learning were investigated in structural modelling; first, the planned investigations of factors associated with accuracy in the delayed Judgment of Learning - memory, executive and affective. Secondly, the contribution of pre-delay JoL as a possible (incorrect) cue used to estimate the delayed JoL. Finally, given the supposed differences between pre and post-delay JoLs, the factors contributing to pre-delay JoL were also investigated. Figure 7.6 summarises the model, now respecified to account for the results presented in Chapter 5.

The main theoretically drive for this relates to the proposal that each judgment is based on differential monitoring of memory stores, as suggested by the *Monitoring dual Memory* hypothesis (Nelson & Dunlosky, 1991). In terms of causal effects, a contributory (directional path) effect is investigated, instead of a covariance, between pre-delay JoL and delayed JoL, because in the majority of the literature there is no proposed correlation between the two, given that underconfidence is typically extinguished after delay. The parameter expectations are outlined and the specified model is summarised below:

1. *Memory* will be associated with both pre- and post-delay Judgments of Learning, with accuracy (higher scores) in pre-delay being positively associated with memory, because predictions will be lower than performance due to the underconfidence with practice effect. In the delayed Judgment of Learning, with maintained underconfidence, the same relationship is expected.

2. *Mood* may be associated with pre-delay Judgment of Learning because, in the presence of mood disorder, it could contribute to underconfidence in future performance.

3. *Executive Function* will be a non-significant path in both Judgments of Learning, supporting a view that mnemonic factors relate to this judgment.

4. As a result of a maintained underconfidence with practice effect, pre-delay Judgment of Learning, will be positively associated with delayed Judgment of Learning, supporting a proposition that the earlier judgment, may be used, inappropriately, as a heuristic in setting the delayed prediction of performance.

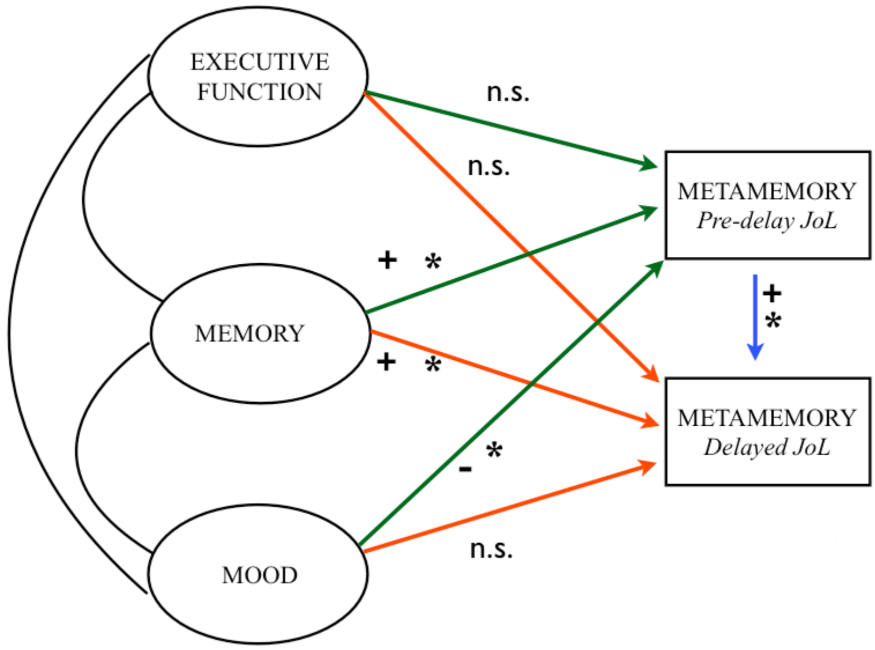


Figure 7.6. Summary of revised *delayed Judgment of Learning* Model, to account for maintenance of underconfidence at delay.

7.4.2.2. Model Testing

Skew, kurtosis and multivariate normality were all within acceptable values. The global model fit statistics were: x^2 = 30.972, df = 27, p = 0.272, CMIN/df = 1.147, GFI = 0.937, CFI = 0.949, RMSEA = 0.039, SRMR = 0.0683. The model is presented in Figure 7.7.

The model explained 18% of the variance in pre delay JoL, and 56% of that for the delayed JoL. For pre-delay JoL, *Mood* and *Memory* were significant parameters: *Mood* B = -3.29, p = 0.043, *Memory* B = 0.542 p = 0.033, but *Executive Function* was not (B = -0.420, p = 0.107). The model suggests that for Pre-delay JoL, higher levels of reported depression were associated with more underconfidence (wider discrepancy). Memory ability was positively associated with accuracy, suggesting the most inaccurate judgments in the pre-delay JoL were associated with greater depression and greater memory impairment.

For the delayed JoL, only *Memory* and the pre-delay JoL were predictors (*Memory*: B = -0.846, p < 0.001; *pre-delay JoL* B = 0.387, p<0.001). Against the expectations of the a priori model, memory ability was negatively related to the *Delayed JoL*. This negative association between memory ability and accuracy suggests better rememberers had wider discrepancies between predicted and actual performance. Executive function was not a reliable predictor in either judgment. Finally, *Pre-delay JoL* was positively associated with this *Delayed JoL*, supporting a contention that participants may have used the previous judgment to inform subsequent judgments. The full statistical outputs for this model are presented in Appendix N.

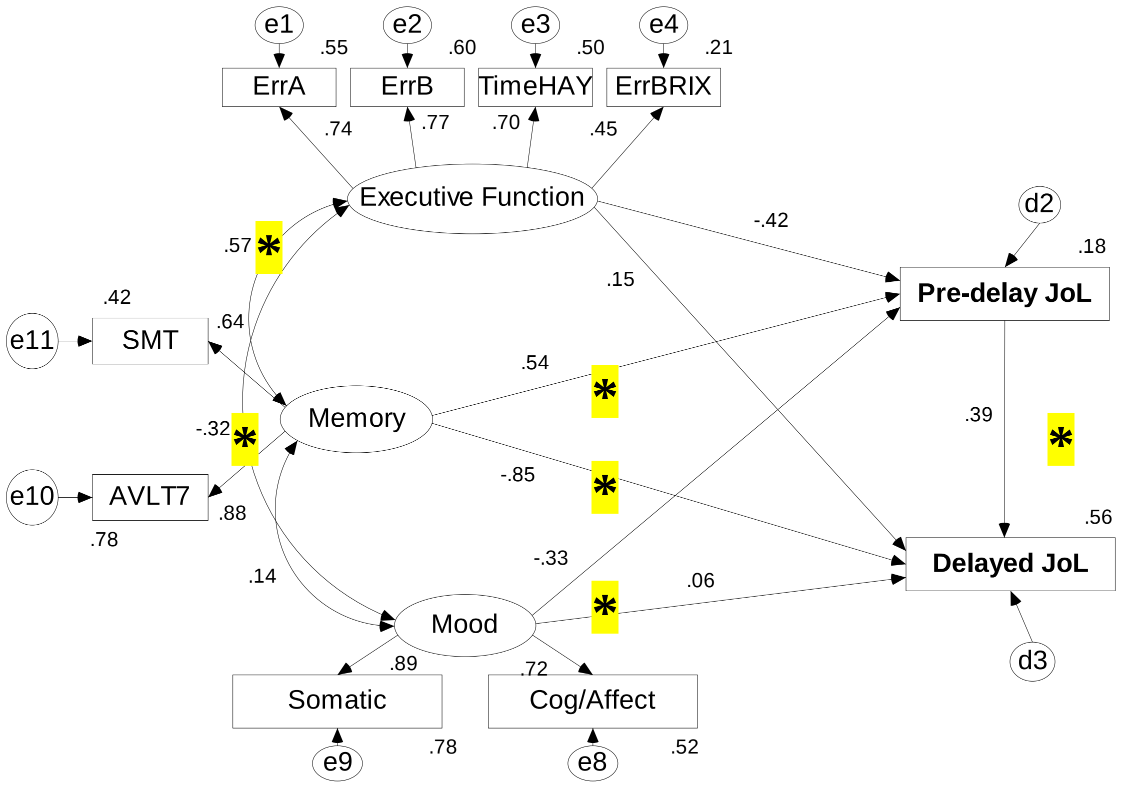


Figure 7.7 Structural model of delayed Judgment of Learning (delayed JoL) * = significant path at $p < 0.05$. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score).e = residuals for observed items, and d = residuals (disturbance) for endogenous latent variables.

7.3.2.3. Judgment of Learning Mediation Model

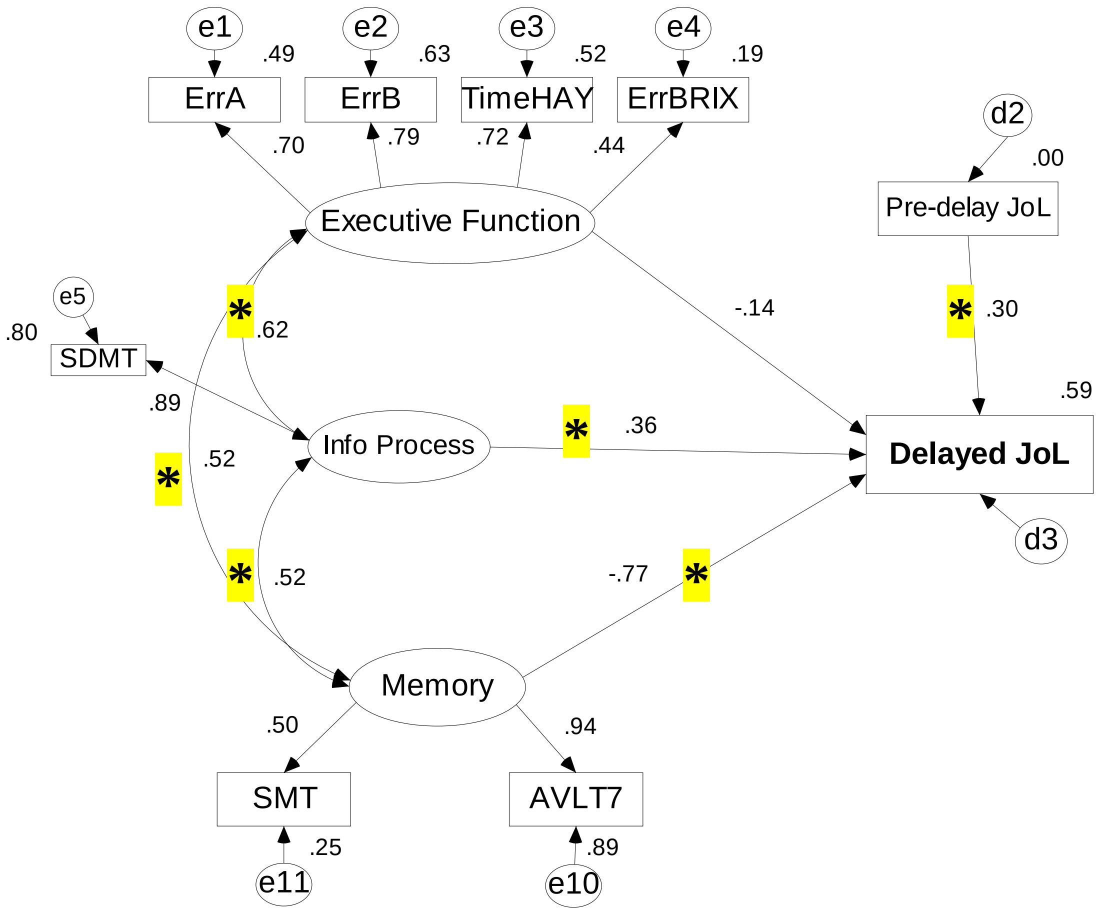
The second Judgment of Learning model investigates whether *Information Processing* might mediate some of the cognitive processing in making the delayed JoL. This is based on the proposal that interrogation of memory store is a primary method for making this judgment, at delay, and that better information processing might facilitate this. For the purposes of this investigation, information processing was investigated as a mediator to *Memory*. This reflects the findings of Kimball & Metcalfe (2003) that the judgment may relate strongly to mnemonic phenomena, and perhaps less of a meta-mnemonic one. Here the proposal is that reduced Information Processing speed might limit the covert retrieval processes, so reducing accuracy. This might occur if a person is unable to make the covert retrieval attempts required to best judge learning after delay. In this sense, despite the increase in learning, which might occur over the 5 learning trials, covert retrieval processes should allow for prediction to reflect that.

7.4.4.1. Model Specification

The JoL model was simplified to focus on the relationships of interest - with the *Delayed JoL* being the single criterion variable. Mood was removed from the model because it was not associated with the delayed JoL. Information processing was assessed for a mediational role in the relationship between *Memory* and *Delayed JoL*. The first analysis assessed a direct effects model of the variables of interest: *Memory, Information Processing, Executive Function, Pre-delay JoL and Delayed JoL.*

7.4.4.2. Model Testing Results

For the direct effects model, fit statistics were generally acceptable with: x^2 = 32.547, df = 23, p = 0.089, CMIN/df = 1.415, GFI = 0.927, CFI = 0.870, RMSEA = 0.065, (90% CI = 0.00 - 0.112) pclose = 0.295, SRMR = 0.0858. In all 59% of the variance in the JoL discrepancy was explained. As before, *Memory* was a reliable contributor (*B* = -0.770, $p < 0.001$) with better recall performance being associated with greater underestimation of performance. *Information processing* ability was a predictor of accuracy (*B* = 0.559, $p < 0.001$) and previous JoL positively contributed to the delayed JoL (*B* = 0.302 $p < 0.001$). The model is shown in Figure 7.8 and the full statistical output for this model is presented in Appendix O



7.8. Structural model establishing the potential for mediation, including Information Processing as a direct effect * = significant path at p < 0.05. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; SDMT = Symbol Digit Modalities Test; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors.e = residuals for observed items, and d = residuals (disturbance) for endogenous latent variables. All latent variable loadings were significant also.

As *Information Processing* was associated (correlated) to a significant extent with *Memory,* and *Delayed JoL,* conditions for assessing mediation of *Memory* were satisfied. Theoretical specification was that *Memory* is the primary factor underpinning Judgment of Learning, as demonstrated in the last model. Based on the MS literature, object-level processes, such as *Memory* may be limited by *Information Processing*. The task in delayed Judgment of Learning is to estimate future recall, based on an assessment of what is known in long term stores (Nelson & Dunlosky, 1991). This is proposed to be carried out using covert retrieval, which typically abolishes underconfidence at delay (Dunlosky & Bjork, 2008).

The direct effects model shown above suggests that accuracy in the judgment is contributed to by *Information processing*, but decreased by *Memory* performance. This may support an understanding of accuracy in this judgment being associated with sufficient processing speed abilities to support the covert retrieval process in conjunction with better memory performance widening the gap between prediction and accuracy. Apparently the two domains work in opposing directions in contributing to accuracy. It is against this background that the mediation of *Memory* by *Information Processing* was tested.

7.3.2.4. Mediation Model Testing Results

The model generated poor fit statistics: x^2 = 39.658, df = 24, p = 0.023, CMIN/df = 1.652, GFI = 0.911, CFI = 0.786, RMSEA = 0.081, (90% CI = 0.030 - 0.125 pclose = 0.128). This offers no support for what might be described as an object-level mediation model, where *Memory* is mediated by processing speed, to support covert retrieval processes and abolish underconfidence at delay.

7.3.2.5. Summary of Judgment of Learning Models.

Direct effects modelling suggested that better recall was associated with greater accuracy in the pre-delay Judgment of Learning, but less accuracy in the delayed judgment. The other factor related to delayed Judgment of Learning was the pre-delay judgment, in line with proposals of Finn & Metcalfe (2007) that it may inform the delayed judgment. Here, this may relate to the maintenance of the underconfidence with practice effect at delay. While mediation was not supported, there was a conflicting contribution of *Information*

Processing and *Memory* to accuracy in the judgment. Information processing speed was reliable associated with accuracy at delay, but better memory performance associated with least accuracy. This finding will be considered further in the following chapter. An interpretation of the finding is that both relationships with delayed Judgment of Learning accuracy indicate better cognitive performance, with information processing explaining the better learning and resultant memory performance the greater disparity between predicted and actual performance. In this regard, neither seem to offer an explanation of the low predictions, instead just the better *actual* performance.

7.3.3. Retrospective Confidence Judgments (RCJ).

Confidence judgements were made for each of the retrieved answers at the cued-recall phase of the Sentence Memory Test. This consisted of participants seeing each incomplete sentence and being asked to supply an answer (including ‘don’t know’) and then giving a confidence in the answer supplied. ‘Don’t know’ responses were rated as no confidence ratings.

Absolute accuracy judgments were derived, based on the proportion (in percentages) of High Confidence correct (accuracy) and percent Low Confidence correct (inaccuracy). The second accuracy measure was RCJ gamma; as a measure of relative accuracy (resolution), it is often reported to be quite low in studies, perhaps because general knowledge is often used as the basis of testing (Dunlosky & Metcalfe, 2009). Here it was high, perhaps relating to the episodic measure used (mean gamma = 0.894 SD = 0.197). Models were tested for both absolute and relative accuracy.

A number of considerations in modelling these judgments here relate to the task used; most of the literature reports performance on general knowledge tasks, many using forced report; that is, no option to report a ‘don’t know’. The relevance of memory type in assessing contributions to these judgments is that in the semantic condition people may be able to use information-based processes in coming to an assessment of confidence e.g. perceived domain expertise, or how much related information about an answer they can retrieve (Koriat et al., 2008). In an episodic task such as that used in this study, there will perhaps be less *information* available, so that people may rely more on experience-based processes to decide on their degree of confidence. Since the structure of the test was a set of incomplete sentences, there were opportunities for cue-familiarity processes to be used however, along with other experience-based indicators towards confidence; ease, speed or contents of retrieval (Nelson & Narens, 1990).

As before, executive, mnemonic and affective contributions are first tested, as direct effects, for each accuracy type. Models are then extended to investigate a role for information processing ability.

7.3.3.1. Model Specification - Calibration or Absolute accuracy

The percent correctly recalled for both High and Low Confidence attributions were first modelled; the model to be tested the two aspects of calibration; *RCJ Accuracy* (% High RCJ, Correct) and *RCJ Inaccuracy* (% Low RCJ correct), with the following expectations:

1. *Memory* will be a contributor to RCJ accuracy, given that accuracy in this judgment is proposed to relate to memory experience.
2. Mood will not contribute to RCJ inaccuracy. The study of Fu et al., (2005) did not support such a relationship between depression and confidence. However, this will be confirmed in the model
3. *Executive Function* will contribute to RCJ inaccuracy reflecting difficulty with ascribing accurate confidence to successful recall. A negative parameter is expected indication better executive performance being associated with lower levels of RCJ inaccuracy. Figure 7.9 summarises the model to be tested and expected parameter estimates.

7.5.2 Model Testing Results

The values for *RCJ Inaccuracy* had a kurtosis value of 5.15, within acceptable ranges according to some heuristics (Kline, 2005), and all other values were within acceptable ranges on both kurtosis and skew measures.

Both High and Low confidence were modeled together and global fit indices were acceptable; x^2=30.433, df=28, p= 0.343, x^2/df= 1.087, GFI= 0.939, CFI= 0.958, RMSEA = 0.030 (90% CI = 0.00-0.085; pclose = 0.665) and SRMR= 0.0751. Full statistical output is available in Appendix P.

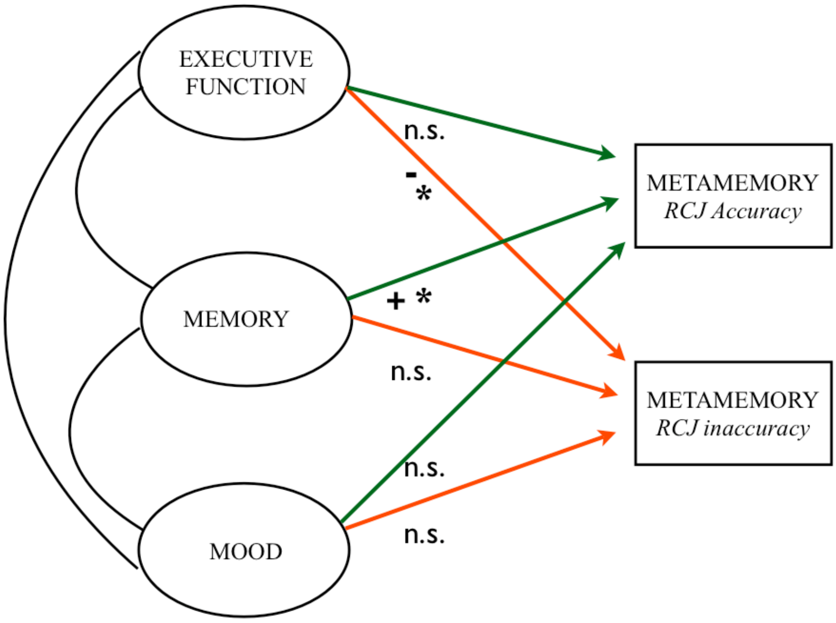
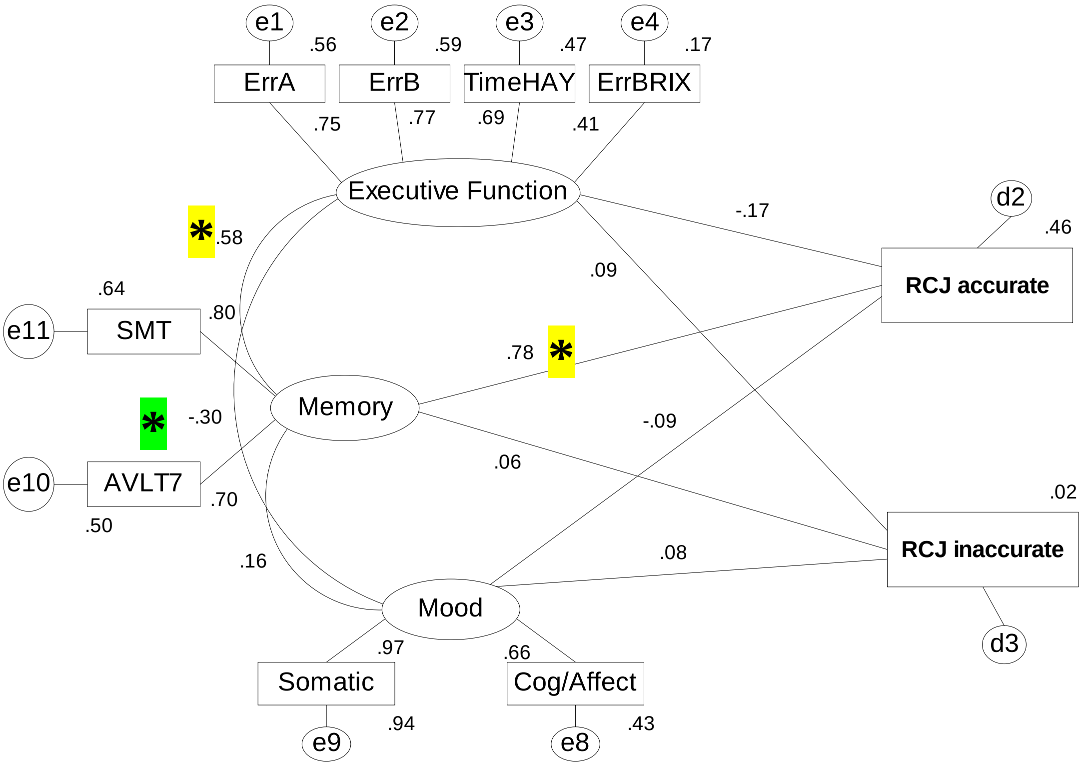


Figure 7.9. Summary of expected parameter and model results for Retrospective Confidence Judgment, absolute accuracy and inaccuracy.

The model, shown in figure 7.10, indicates the significant parameter for *Memory* on RCJ Accuracy (B = 0.780, p = 0.001), with 46% of the variance explained. None of the included predictors - *Memory, Mood* or *Executive Function* - contributed to a significant extent to explaining variance in RCJ inaccuracy. Only *Executive Function* and *Memory* correlated to a reliable extent (r = 0.585; p = 0.023), with *Mood* and *Executive Function* correlated r = -0.30, p = 0.061. In both the RCJ accuracy and RCJ inaccuracy indices actual responses were correct, so parameter estimates relate to the assessment of confidence, in the context of availability of the correct sentence completion. High confidence appeared based on memory related experience, but not other inferential judgments, whereas Low confidence did not relate to inferential or mnemonic experiences. *Mood* was not a reliable predictor in either case.



7.10 Structural model testing factors contributing to Retrospective Confidence judgment, absolute accuracy (RCJ accurate) and inaccuracy (RCJ inaccurate). * = significant path at p < 0.05. * = significant path at p < 0.10. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score).e = residuals for observed items, and d = residuals (disturbance) for endogenous latent variables. All latent variable loadings were significant also.

7.3.3.3. Model Specification - Relative accuracy

The second model investigates the relative accuracy of the Retrospective Confidence Judgment. This is indicated by the RCJ gamma score, and indexes the ability to ascribe confidence judgments on one item, relative to another. The literature generally presents findings that support memory-related processes (ease of retrieval, amount retrieved) being linked to this judgment accuracy under free report (Koriat et al., 2008). As this was the structure here, the expectations for this model are for *Memory* to be positively associated with accuracy. Figure 7.11 is the a priori model to be tested.

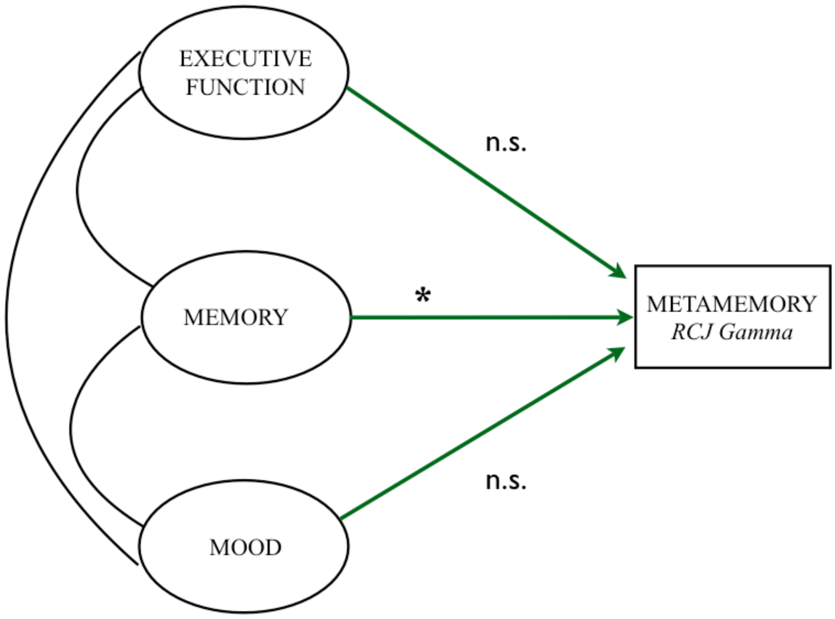


Figure 7.11. The a priori structural model for Retrospective Confidence gamma.

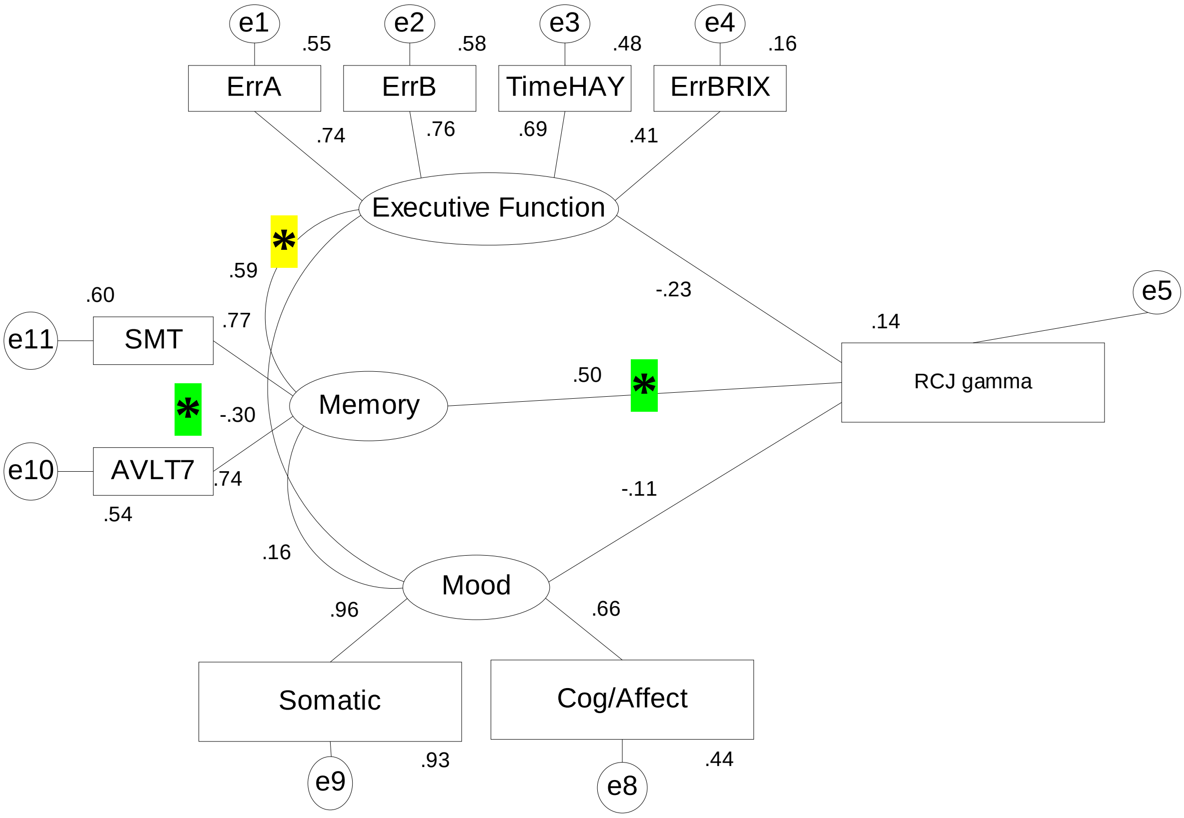
7.3.3.4. Model Testing

All but one gamma value was ≥ 0, so interpretation of the gamma statistic (a correlation) is of larger values indicating greater accuracy. Initial modelling, in spite of significant non-normality in the distribution of gamma scores, was carried out. The model testing suggested no statistically significant relationships between the exogenous variables and RCJ accuracy, though memory approaches reliability as an indicator; *Memory* ($p = 0.068$), *Mood* ($p = 0.458$) or *Executive Function* ($p = 0.336$).

The model explained 14% of the variance in RCJ Gamma. *Memory* and *Executive Function* were correlated (r = 0.594, p = 0.019) and *Mood* and *Execuitve Function* again almost reached significance at the p <0.05 level (r = -0.301, p = 0.061). The global fit statistics were: x^2 = 27.975 df = 22 p = 0.177, CMIN/df = 1.272 CFI = 0.937, GFI = 0.937, CFI = 0.905 and RMSEA = 0.052 (90% CI = 0.00 - 0.104; pclose = 0.436), SRMR = 0.0698 suggesting acceptable global fit. Figure 7.12 shows the model, with full statistical output provided in Appendix Q.

There was significant kurtosis in the gamma distribution, with most values having a high RCJ gamma score (mean = 0.894, SD = 0.197; Kurtosis critical ratio = 33.45). It seems likely that the distribution of data limited the model's interpretative power. In order to retest the model based on more acceptable distributional characteristics, a number of transformations were carried out on the data, including reflecting followed by Square-root transformation, Log10 and log e, aimed at reducing the higher values and extending the range of scores. These transformations did not improve normality, or the outcome of modelling, and are not reported.

As a speculative execise, outlying values were deleted, based on a search for both extreme and outlying cases, and included eight cases. The model was retested and resulted in 25% of the variance in the RCJ gamma being explained, and an significant parameter estimate for Memory (B = 0.540, p = 0.029) with fit statistics: x^2 = 31.009 df = 23 p = 0.100, CMIN/df = 1.392, GFI = 0.867, CFI = 0.922, and RMSEA = 0.066 (90% CI; 0.00 - 0.099; pclose = 0.296) SRMR = 0.0849. The model is proposed as indicative rather than demonstrative of expected findings, given the distributional properties of the data, absolute accuracy results for the RCJ process, and substantive theory about how RCJ judgments, under free report conditions, appear to be supported by retrieval-related experiences; amount retrieved, content retrieved, speed and ease of retrieval (Koriat et al., 2008).



7.12 Structural model testing factors contributing to Retrospective Confidence judgment gamma, or relative accuracy. * = significant path at $p < 0.05$. * = significant path at $p < 0.10$. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items. All latent variable loadings were significant also.

7.3.3.5. Mediation models for Retrospective Confidence Judgments

A direct effects model, using only the RCJ absolute accuracy measure (percent High Confidence, and correct), with the inclusion of *Information Processing* did not establish a potential for mediation, and *Memory* remained the only significant parameter (B = 0.790, p = 0.001). Correlations were moderate to high between *Memory* and *Executive Function*, and *Executive Function* and *Information Processing* (both r = 0.607). Given the non-significant parameters for both *Mood* and *Information Processing* in relation to RCJ confidence, support for a mediated process is lacking. This would support suggestions that Retrospective Confidence Judgments, where accurate, are in large part supported by experience-based mnemonic cues, rather than the potentially more effortful information or inference based judgments (Koriat et al, 2008).

For the RCJ gamma model, there was, again, no direct association with the proposed mediator, *Information Processing,* so no mediation was tested. This may have occurred for reasons of RCJ related contributions not typically being related to effortful or controlled processing, but could also have been because of the distribution of the gamma scores.

7.3.3.6. Summary of Retrospective Confidence Models

Based on the direct effects model of absolute accuracy (calibration), the strong association of memory ability and accuracy in the RCJ task likely relates to memory-experience indicators. The impact of information processing speed was not expected to contribute as these memory experiences are proposed to be ‘parasitic’ on the retrieval process itself (Koriat et al., 2008:120). In this respect it offers some support for these retrieval-based experiences not being impacted by information processing capacities, perhaps because of their procedural quality (Reder & Schunn, 1996)

In sum, only concordant responses - high confidence in answers that were correct, presented expected relationships, namely the proportion of answers given a high confidence (and were correct), being positively associated with memory performance. Of interest too is the lack of any relationships between presumed predictors and the inaccuracy condition, the proportion of items for which a Low confidence rating was given, which were actually correct. The findings from relative accuracy (RCJ gamma) to some extent mirror the absolute accuracy findings, with memory ability alone relating to accuracy.

7.3.4. Feeling of Knowing judgment.

Feeling of Knowing judgments (FoKs) are typically assessed in respect of semantic memory, rather than newly learned, episodic, information; suggested mechanisms for achieving accuracy include domain-familiarity, cue-familiarity and target accessibility (Dunlosky & Metcalfe, 2009), but not direct access to the target item. FoK items in this study were based on sentence completions for which recall already failed, and episodic rather than semantic recall.

In addition to the cue-familiarity information, accessibility information may be available, based on failed retrieval attempts, giving mnemonic information from which an inference about recognition might be drawn (Koriat, 1993; Koriat et al, 2008). Related to Beatty & Monson's (1991) findings that, for a MS sample, both memory and executive impairment was associated with increasing inaccuracy in FoK, it is proposed that both executive function and memory will be relevant predictors. Additionally, to address the relevance of mood in task-based judgment, this will also be included as a predictor of performance in the models. As before, accuracy in both calibration and resolution was investigated.

7.3.3.1. Model Specification - Absolute Accuracy

The model investigates differences in performance between those items for which a High FoK was ascribed, and those for which a Low FoK was ascribed. This difference is not about association (relative accuracy), but rather proportion, measured here as percent correct for each of *High FoK*, or strong Feelings of Knowing for an answer (measured in the recognition trial) and *Low FoK*, or weak Feelings of Knowing. The two measures are used in respect of correct recognition. The endogenous variables in the model therefore consist of the proportion of High FoK with correct recognition, and the proportion of Low FoKs with correct recognition; both are measured in percentages.

The model assumes a greater degree of explicit and controlled processing, in addition to implicit experience-based processes, that the previous models, mainly because retrieval has already failed; FoKs are made only on unrecalled or incorrectly recalled Sentence Memory Test items. It is proposed that a range of distal cues may be used to make inferences about future recognition. The aims were to assess the contributions of memory and executive function to investigate inferential and experienced-based processes on FoK accuracy and inaccuracy, and expectations of this model are listed below with parameter projections outlined in figure 7.13

1. Expectations were that *Memory* and *Executive Function* were both associated positively and significantly with accuracy (High FoK, correctly recognised), and *Executive Function* negatively with inaccuracy (Low FoK, correctly recognised). The negative association with FoK inaccuracy was proposed as a process in which frontal memory contributions were compromised. *Memory* is proposed to remain positively associated with FoK inaccuracy, because it is based on items for which there was correct recognition.

2. *Mood* is assessed for a contribution, but was expected not to provide a significant parameter for High FoK, correctly recognised. For Low FoK, correctly recognised. It was considered as a potential contributor, explaining why Low FoKs might be ascribed to recognisable targets as being a function of affect-associated underconfidence.

3. As many have suggested that FoK judgments, especially for new learning, appear to be effortful, inclusion of information processing as a mediator to the inferential (executive) and mnemonic components is modelled subsequently. This is different to previous models, which assumed that having available memory-experiences would not constitute an effortful process.

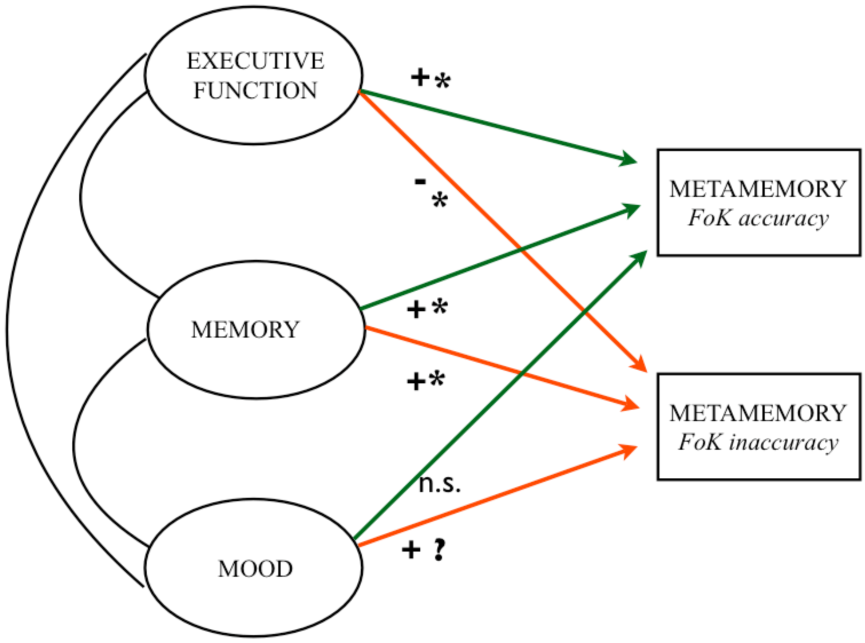


Figure 7.13. The a priori model of Feeling of Knowing (FoK) calibration. FoK accuracy is measured by proportion of High FoKs that were recognised and the FoK inaccuracy as the proportion of Low FoKs subsequently recognised. n.s. = expected non-significant parameter. * = expected significant parameter and ? = an unknown contribution to be investigated in the model + and - indicate the expected direction of contribution.

7.3.4.2. Model Testing

Univariate skew and kurtosis, and multivariate normality assumptions were generally met. Fit statistics were: x^2 = 35.482, df = 28, p= 0.156, x^2/df = 1.276, GFI = 0.928, CFI = 0.885, RMSEA = 0.052 (90% CI 0.00 - 0.099, pclose = 0.442) and SRMR = 0.0729. 26% of the variance of accuracy and 38% of inaccuracy was explained in the model, which is shown in Figure 7.14.

For accuracy in this task, measured by % High FoK correct, parameter estimates were significant for *Memory* (B = 0.505, p = 0.032) but not for *Mood* (p = 0.623) or *Executive Function* (p = 0.834). Better recall was associated with higher proportions of concordance between High Feelings of Knowing and subsequent correct recognitions; suggesting that weak or content-less mnemonic cues, might have assisted the judgment in this accuracy condition.

For inaccuracy, indicated by the proportion of Low Feelings of Knowing subsequently correctly recognised, (% Low FoK correct), both *Memory* and *Executive Function* were significant contributors (B = 0.670, p = 0.002 and B = -0.469, p = 0.036 respectively); *Mood* was not a contributor (B = 0.062, p = 0.717). The relationship in the inaccuracy condition was positive for *Memory*, but negative for *Executive Function*, suggesting those with better executive abilities had less had less inaccuracy.

Noteworthy is that *Executive Function* only approached a significant correlation with memory in this model (r = 0.540, p = 0.069) perhaps suggesting some discrimination in the respective contribution of each to the metamemory measures, rather than a general higher order association with ‘Cognition’. In all other models there were robust correlations between the two. *Mood* correlated with *Executive Function* (r = -0.313, p = 0.022). Full statistical outputs are presented in Appendix R.

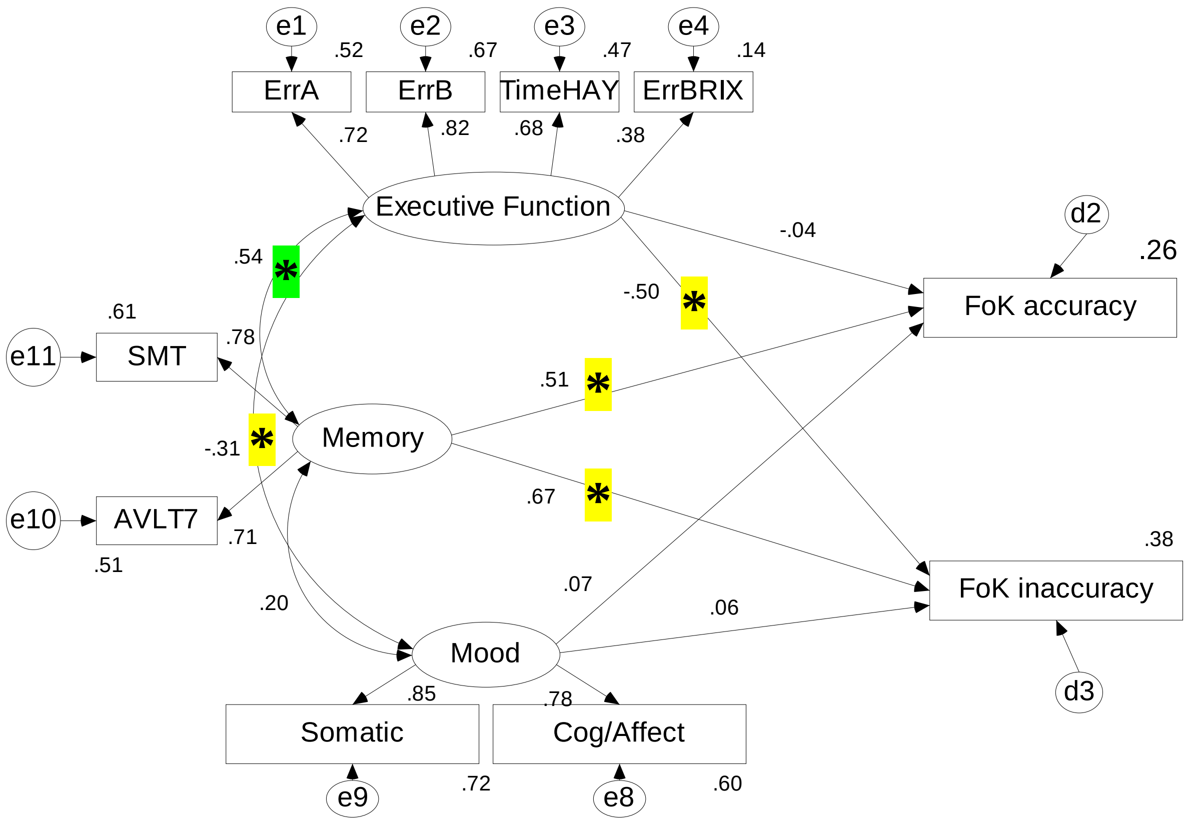


Figure 7.14. Structural model testing factors contribution to Feeleing of Knowing calibration, or absolute accuracy. * = significant path at $p < 0.05$. * = significant path at $p < 0.10$. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items. All latent variable loadings were significant also. FoK accuracy is measured by proportion of High FoKs that were recognised and the FoK inaccuracy as the proportion of Low FoKs subsequently recognised

In order to test for any relationship between accuracy and inaccuracy in this FoK measure, a model in which the two were freed to correlate was tested. The result was r = -0.002, p = 0.989 suggesting they were unrelated when the contributory factors were taken into account, perhaps supporting a view that inaccuracy is not simply a failure of accuracy, but relates to a separate set of processes.

The differential contribution of executive abilities - no direct association in the accuracy condition, negative association in the inaccurate condition - fits with a proposal that inaccuracy in neurological samples is associated with executive deficit (Beatty & Monson, 1991; Pannu & Kaszniak, 2005). Inaccuracy is increased in the presence of executive deficit - the ascribing of Low FoKs, in the context of subsequent correct recognition - suggesting that impaired inferential mechanisms might override mnemonic availability cues in this condition. This possibility was pursued for a statistical assessment by testing a model in which *Executive Function* mediated *Memory* in the FoK inaccuracy condition.

7.3.4.3. Mediation Model for Feeling of Knowing Inaccuracy.

Mood was excluded as a predictor variable in the model for mediation, as it was demonstrated not to relate to either endogenous variable. As before, with a requirement for a mediator to be tested, it must relate to the endogenous variable of interest - the FoK inaccuracy measure. From the direct effects model presented above, both *Executive Function* and *Memory* were contributory factors. Given that an initial FoK may happen faster than a retrieval attempt, there may be a two-step process in completing the FoK judgment (Miner & Reder, 1994; Koriat & Levy-Sadot, 2001). In specifying such a model of mediation, the main question relates to the likely processes.

For the Low FoK condition, of interest here, any mnemonic contributions to forming the judgment might be discounted in some way by other inferential information - this would support *Memory* being mediated, with negative effect, by *Executive Function* in the FoK inaccuracy condition. An initial experience-based process might thus be discounted by more explicit, but faulty, information-based judgment, undermining potentially useful mnemonic experience (Koriat & Levi-Sadot, 2001; Koriat, 2008; Dunlosky & Metcalfe, 2009).

Such a process might be expected in people with more executive dysfunction, which has been linked with memory related deficits in source failure in remembering, and confabulation (Johnson et al., 2000). Both disorders have been associated with insufficient regard to accuracy, typified, by errors in attributing appropriate weight to memory-related diagnostic information (Johnson et al., 2000), or over reliance on heuristic as opposed to systematic processing which might better note useful mnemonic cues and use them to infer likelihood of future recognition (Johnson et al., 2000; Moscovitch & Wincour, 2002). The model presented (figure 7.15) tested whether *Executive Function* might mediate *Memory* in FoK inaccuracy performance.

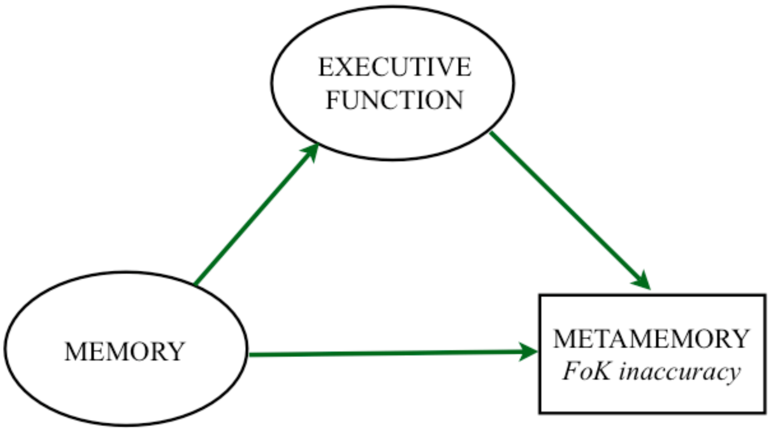


Figure 7.16 Schematic model of mediation of Memory by Executive Function in leading to Feeling of Knowing calibration inaccuracy. FoK inaccuracy as the proportion of Low FoKs subsequently recognised.

Model fit statistics were: x^2 = 19.439, df = 12, p = 0.078 x^2/df = 1.620, GFI = 0.944, CFI = 0.833, RMSEA = 0.079 (90% CI = 0.00 - 0.141; pclose = 0.209) SRMR = 0.090. Sobel test z = -2.105, p = 0.035 (MacKinnon et al., (2002), adjusted = p < 0.01). Parameter estimates were significant for all mediational paths. The model fit indices and Sobel test are consistent with a mediation view of the relationship between *Memory* and *Executive Function.* Figure 7.16 presents the model results and Appendix S the full statistical output.

This model supports a proposition that the direct effect of memory is reliably different when executive function is included as a mediator of the *Memory - FoK Inaccuracy* relationship. The model has limited utility in that it presents the same fit statistics, x^2 and degrees of freedom as a model in which both *Memory* and *Executive Function* were correlated direct effects, so that statistical comparison between the two models is not possible. Its benefit is proposed in addressing the nature of the relationship between the two predictors, as an alternative to a correlated direct effects model. As such, it presents the same associations, but interpreted in terms of causal, as opposed to correlated, relationships between predictors, but with some theoretical support (Johnson et al., 2000)

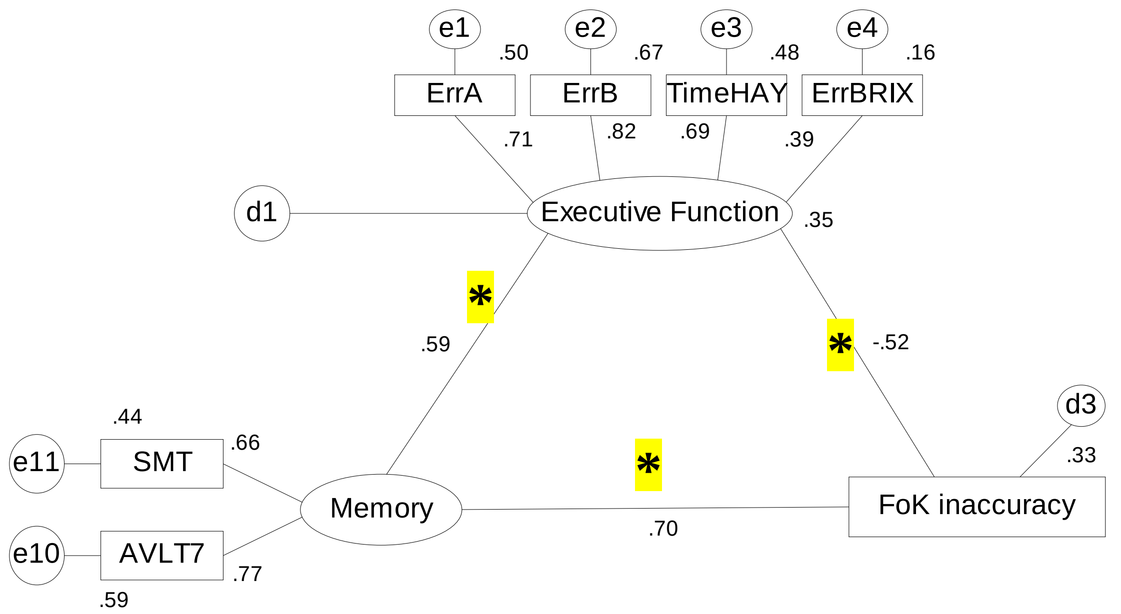


Figure 7.16. Structural mediation model of *memory* mediated by *Executive Function* in Feeleing of Knowing inaccuracy. * = significant path at $p < 0.05$. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. e = residuals for observed items. All latent variable loadings were significant also. FoK inaccuracy is the proportion of Low FoKs, subsequently recognised.

7.3.4.5. Information Processing mediation in FoK Calibration

The main additional aim in FoK model testing was to investigate whether *Information Processing* might mediate the processing involved in metamemory accuracy. In a direct effects model of FoK calibration, with all four predictors, *Information Processing*, *Memory*, *Executive Function* and *Mood* included, *Information Processing* did not provide a significant relationship with either of the FoK measures (for FoK accuracy, B = -0.215, p = 0.198; for FoK inaccuracy B = -0.063, p = 0.704) and so could not be assessed for mediation, despite correlating with both *Memory* and *Executive Function*.

These negative results in terms of information processing are worth considering from the perspective of effortful cognition and the proposed role of *Information Processing* in that. The *Memory* and *Executive Function* direct effects model has already demonstrated that only *Memory* was a contributor for absolute accuracy. Measured in this way, it supports a proposal that with each incomplete sentence acting as a cue to diagnose future recallability, cue-familiarity may be what supports the accurate calibration decision, and that deliberative or controlled processing contributions may not be used to a significant extent. Given that both absolute accuracy and absolute inaccuracy were assessed on the basis of subsequently recognised items, familiarity may have been the key contribution.

7.3.4.6. Model Specification - Relative Accuracy in Feeling of Knowing

The aims in this set of models are to test factors associated with accuracy in FoK gamma and more specifically the findings of Beatty & Monson (1991), that in people with MS, increasing inaccuracy in FoK is associated with increasing executive dysfunction, in addition to memory deficits. Additional models investigate a contribution of information processing abilities towards accuracy in FoK judgment, based on the proposal that it can be considered an effortful cognitive judgment process, potentially constrained by the reduced processing ability common in MS.

Specification of FoK gamma direct effects model

As before, the initial model was based on direct effects of *Memory*, *Executive Function* and *Mood* on FoK gamma. Model expectancies are as follows:

1. *Memory* and *Executive Function* will demonstrate a positive contribution towards accuracy (higher FoK gammas), in agreement with the findings of Beatty & Monson (1991).

2. Mood may demonstrate a negative contribution to accuracy, indicating more mood disorder is associated with lower accuracy.

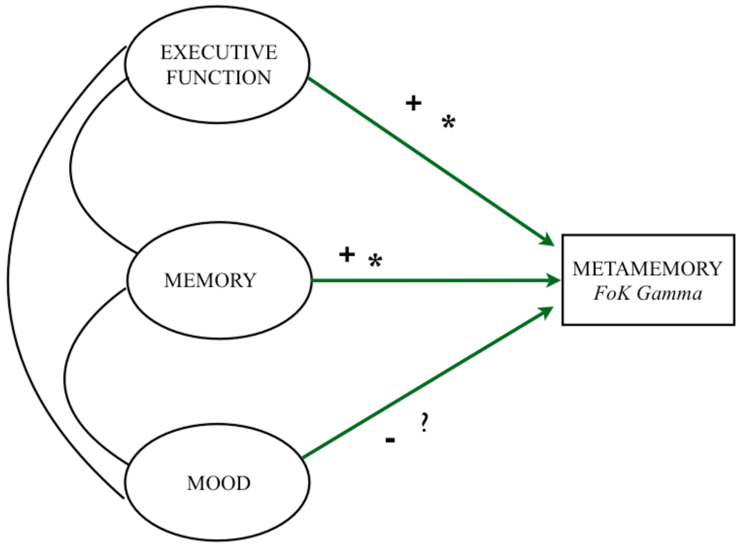


Figure 7.17. The a priori model of Feeling of Knowing (FoK) gamma.* = expected significant parameter and ? = an unknown contribution to be investigated in the model + and - indicate the expected direction of contribution.

7.3.4.7. Model Testing

As encountered previously, a negative error variance was calculated for the error term for the Somatic factor score for *Mood*. Error variance for this item was therefore set to a non-zero value (.01) in line with the previously cited recommendation (Joreskog & Sorbom, 1998). Results for model fitting were: x^2 = 25.483 df = 23 p = 0.326, CMIN/df = 1.108 CFI = 0.961, GFI = 0.943 and RMSEA = 0.033 (0.000 - 0.091; pclose = 0.620), SRMR = 0.0719.

Higher gamma scores indicate greater accuracy. In terms of contribution to FoK gamma, *Mood* was a significant item (B = 0.286; p = 0.052), but neither *Memory* nor *Executive Function* contributed directly (B = 0.103, p = 0.618 and B = 0.139, p = 0.522). The model explained 11% of the variance in the FoK gamma score. *Mood* correlated with *Executive Function* (r = -0.297, p = 0.017) and *Memory* with *Executive Function* (r = 0.588, p = 0.005). Figure 7.18 summarises the model results, full statistical outputs are available in Appendix T.

This model suggests that those scoring higher on a depression measure had higher gamma scores (more accuracy), contradicting expectations. Further, no reliable direct contribution of *Memory* or *Executive Function* was evident. The results were not as expected generally, although general fit of the model was good in terms of global indicators of fit. Reasons for lack of expected findings may relate to the complexity of the task, which has two implications; first that the gamma findings are unreliable, and second the processing demands of the task meant that requirements for effortful cognitive operations are high; without considering a mediational relationship of information processing actual relationships may not be uncovered. This latter prospect was addressed first as it was a planned assessment. It was explored by means of assessing indirect contributions of *Memory* and *Executive Function*, mediated by *Information Processing*.

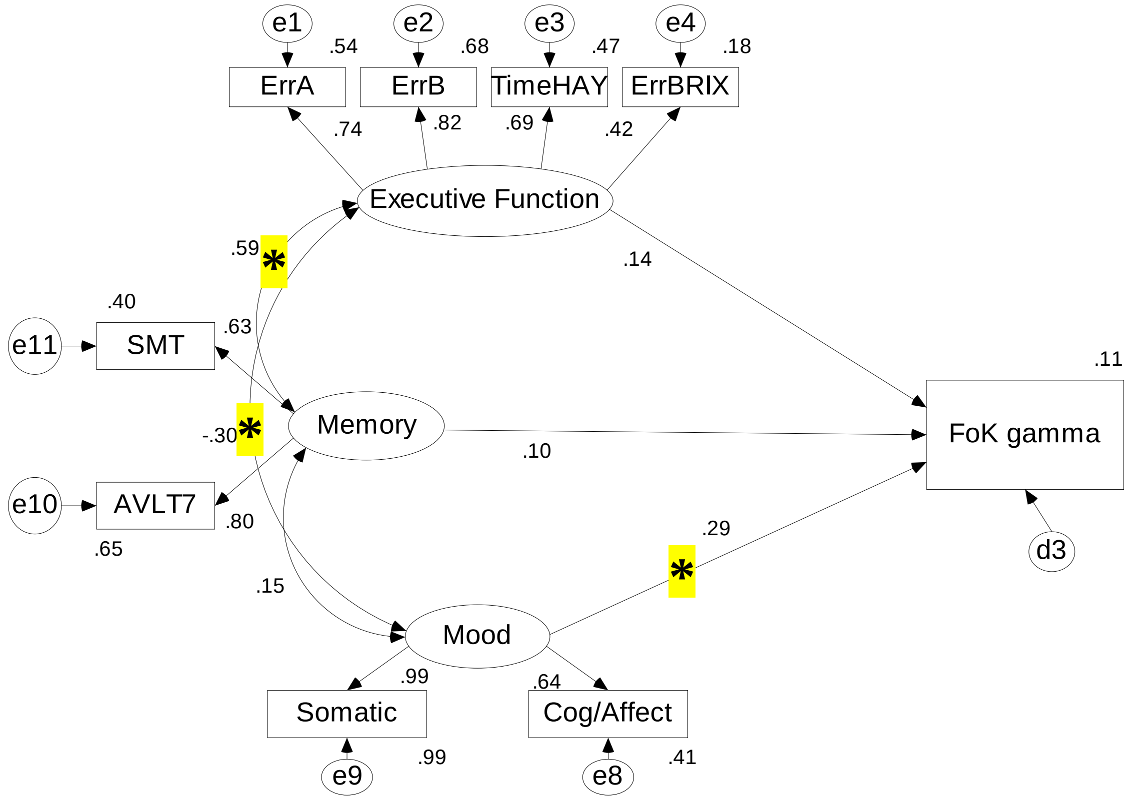


Figure 7.18. Structural model of factors contributing to Feeleing of Knowing accuracy, measured by gamma. * = significant path at p < 0.05. * = significant path at p < 0.10. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items. All latent variable loadings were significant also.

7.3.4.8. FoK gamma Mediation Model Specification

As an initial step, a direct effects model of four correlated latents (*Memory, Executive Function, Information Processing* and *Mood*) was assessed, to establish statistical support for a mediational model, which required a significant relationship between the mediator (*Information Processing*) and the to-be-mediated (*Executive Function* and/or *Memory*), in the presence of significant effect of the mediator on the FoK measure.

7.3.4.9. Model Testing Results

Results of the modelling supported a significant relationship between *Executive Function* and *Information processing* (r = 0.608, p <0.001) and *Memory* and *Information Processing* (r = 0.565, p <0.005). Neither *Mood* and *Information processing,* nor *Mood* and *Memory* correlated (r = -0.122, p = 0.309 and r = 0.126, p = 0.338 respectively). Direct effects were evident for both *Information processing* to FoK gamma (B = -0.344, p = 0.049) and, as before, *Mood* to FoK gamma (B = 0.277, p = 0.049), but not for *Executive Function* and gamma (B = 0.281, p = 0.192) or *Memory* and gamma (B = 0.203, p = 0.312). Global fit of this model, which accounted for 17% of the variance in FoK gamma was acceptable; x^2 = 29.286, df = 28, p = 0.398, x^2/df = 1.046, GFI= 0.941, CFI= 0.980, RMSEA= 0.022 (90% CI 0.00 - 0.082, pclose= 0.714) and SRMR = 0.0712. The model is presented in Figure 7.19, and statistical outputs in Appendix U.

The negative parameter estimate for *Information Processing* to FoK gamma suggests that those with better processing speed tended to have lower gamma values, indicating less accuracy. This finding runs contrary to what was expected from the perspective of information processing abilities supporting the processing involved in making the accurate judgment in the FoK task. This finding will be considered here in how it might guide further model specification.

The a priori mediation model proposed that *Information Processing* mediates the effect of both *Executive Function* and *Memory* on FoKs. Theoretically, this tests the idea that in the FoK task, effortful processing, which may not have available the same experience-based cues as other metamemory judgments, is demanded. Mnemonic and inferential effort is proposed to be high in a task where retrieval has already failed, and where item by item ordering of FoK strength is required; greater reliance on a range of more distal cues may be necessary in promoting accuracy.

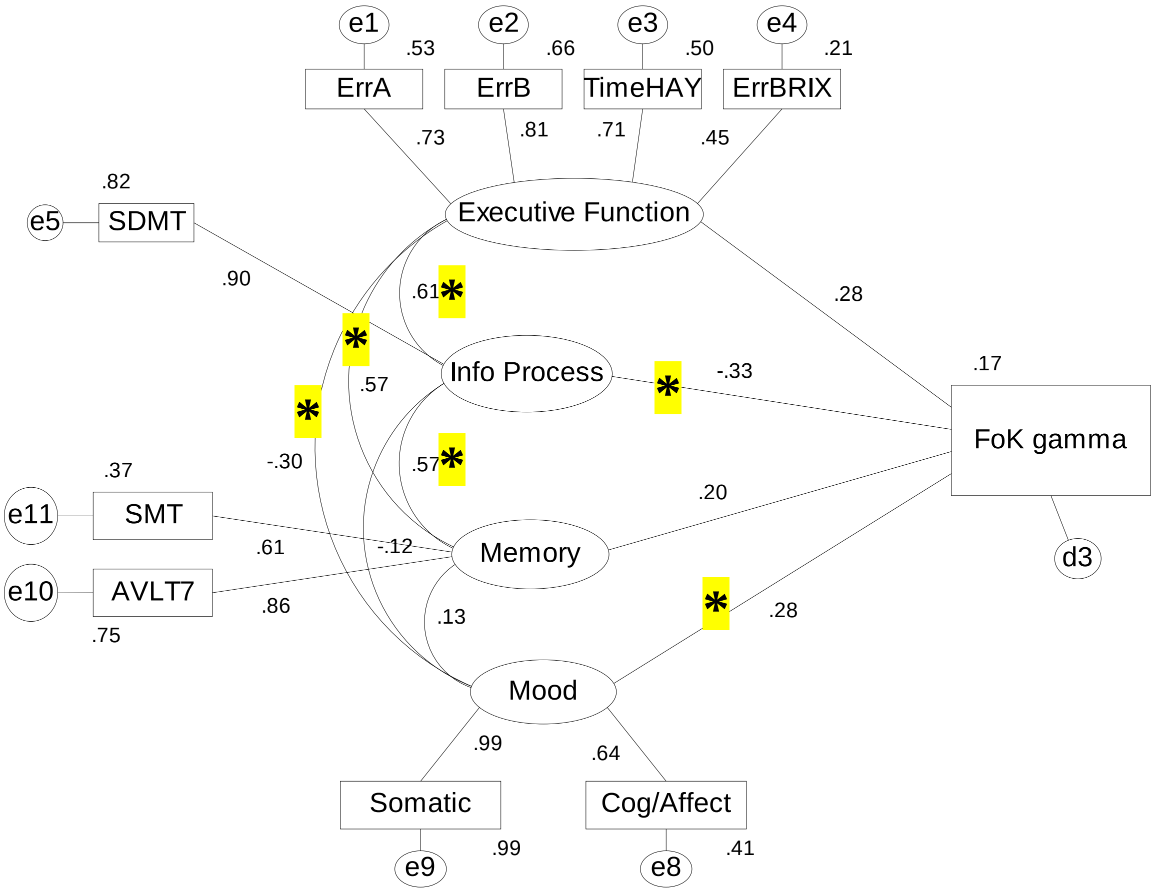


Figure 7.19. Structural model of factors, including Information Processing, contributing to Feeleing of Knowing accuracy, measured by gamma. This model established the potential for mediation. * = significant path at p < 0.05. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. SDMT + Symbol Digit Modlaities Test; Cog/Affect = Cognitive Affective Beck Depression Inventory items (factor score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items. All latent variable loadings were significant also.

In proposing mediation by *Information Processing*, which has demonstrated a negative beta weight, contrary to that expected, but a positive correlation with both *Memory* and *Executive Function* as expected, it may be more appropriate to consider that this indicated a confounding, or suppressing, relationship on the variables of interest - *Memory* and *Executive Function* (MacKinnon, Krull & Lockwood, 2000; Iacobucci, 2008). The lack of a significant path, notable for *Executive Function* as a significant contributor to FoK accuracy, may relate to the impact of processing speed suppressing such a relationship. The bivariate correlation between *Executive Function* and *Information processing* is high at r = 0.61 perhaps indicating some collinearity, or shared variance. If variance relating to information processing is removed by setting it as a mediator, a relationship between *Executive Function* and FoK accuracy, which would be expected from a range of findings in neurological populations (Beatty & Monson, 1991; Shimamura & Squire, 1986; Souchay et al., 2003; Pannu & Kaszniak, 2005; Souchay, 2007), might be demonstrated.

In respect of *Memory*, a similar relationship may be demonstrable; the contribution is less clear, based on more inferential accounts of making FoKs, and because these judgments are being made on already unrecallable targets. The model will investigate the role of *Information Processing* as mediator or confound in this manipulation of the model. As theoretical specification proposes that the relationship of *Memory* and *Executive Function* to FoK relates to a shared supression by information processing speed, the correlation between these two items is modelled through the mediator alone. Some support for this specification is justifiable by the weak correlation of these two latent variables in FoK, compared to other models. Against the background of the parameter sign for *Information Processing*, the revised *a priori* model to be tested is outlined in Figure 7.20 proposing the expected relationships.

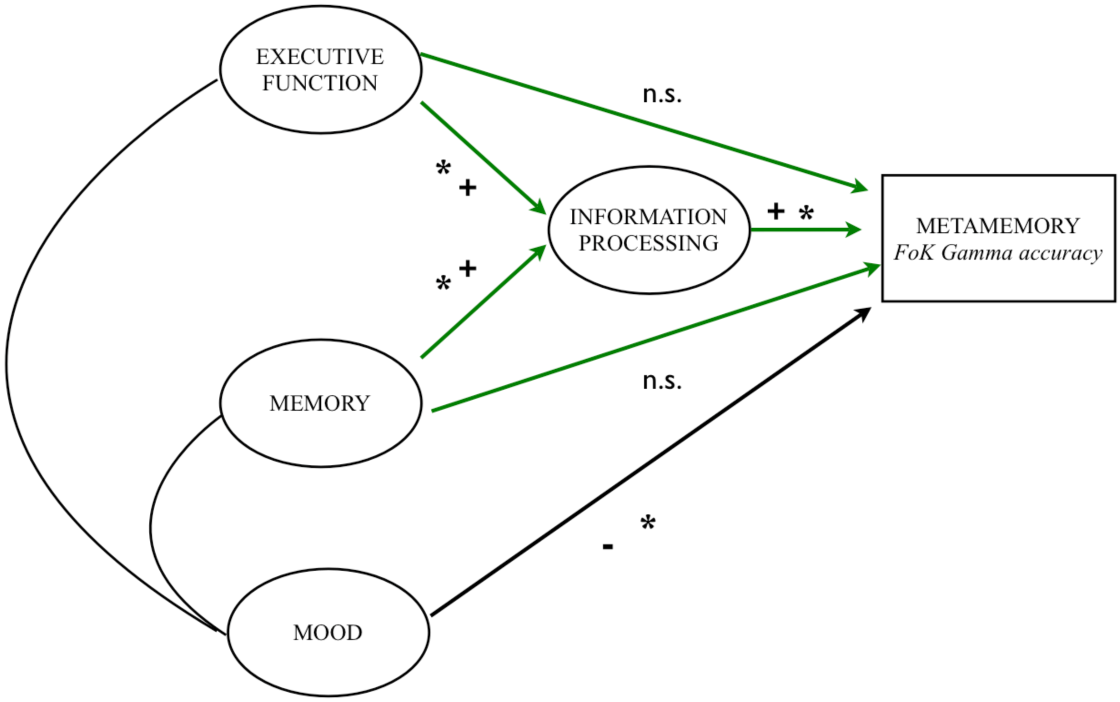


Figure 7.20: The adjusted a priori model of factors contributing to Feeling of Knowing gamma accuracy, with Information processing mediating the relationship between both executive and memory ability in achieving accuracy. Mood disorder has already been demonstrated to relate to accuracy. * Indicates expected significant parameters; n.s. = non-significant parameters and + or - indicates the expected direction of the effect. The two mediational relationships are shown in green.

7.3.4.9. Mediation Model Testing Results

Model fit statistics were generally acceptable; x^2 = 37.786, df = 30, p = 0.155, x^2/df = 1.260, GFI= 0.924, CFI= 0.880, RMSEA= 0.051 (90% CI 0.00 - 0.097, pclose= 0.452) and SRMR = 0.1518. In this model, a significant parameter estimated for *Executive Function* and FoK accuracy was established (B = 0.33, p = 0.020), in addition to near significant (at p <0.05) estimates for *Mood* and *Information Processing* and FoK accuracy (B = 0.25, p = 0.69 and B = -0.27, p = 0.056).

Of note in the model, is support for a direct effect of executive function in increasing accuracy in FoK judgments. Parameter estimates for *Information Processing* are of not in the expected direction, challenging the proposal that information processing speed is positively related to metamemory accuracy; instead it suggests that processing speed may confound the relationship between executive abilities and accuracy, and contribute towards inaccuracy in this measure. Sobel tests of mediation suggested the mediation relationship led to significant changes in the direct relationship between *Executive Function* and FoK gamma as a result of the inclusion of information processing as a mediator (Sobel *z* = -1.257, p = 0.20, adjusted p < 0.05), and in *Memory* (Sobel *z* = -1.273, p = 0.20, adjusted p < 0.05), though the direct effect of Memory was non-significant, even with the mediational relationship. The mediational role of *Information Processing* might therefore best be characterised as one of suppression, against expectations of the a priori model's parameter sign. A final finding of this model was the failure to maintain a significant contribution to the latent variable *Executive Function* by one of the observed variables the Brixton Spatial Anticipation Test. The full set of statistical outputs in available in Appendix V, and the model is shown in figure 7.21.

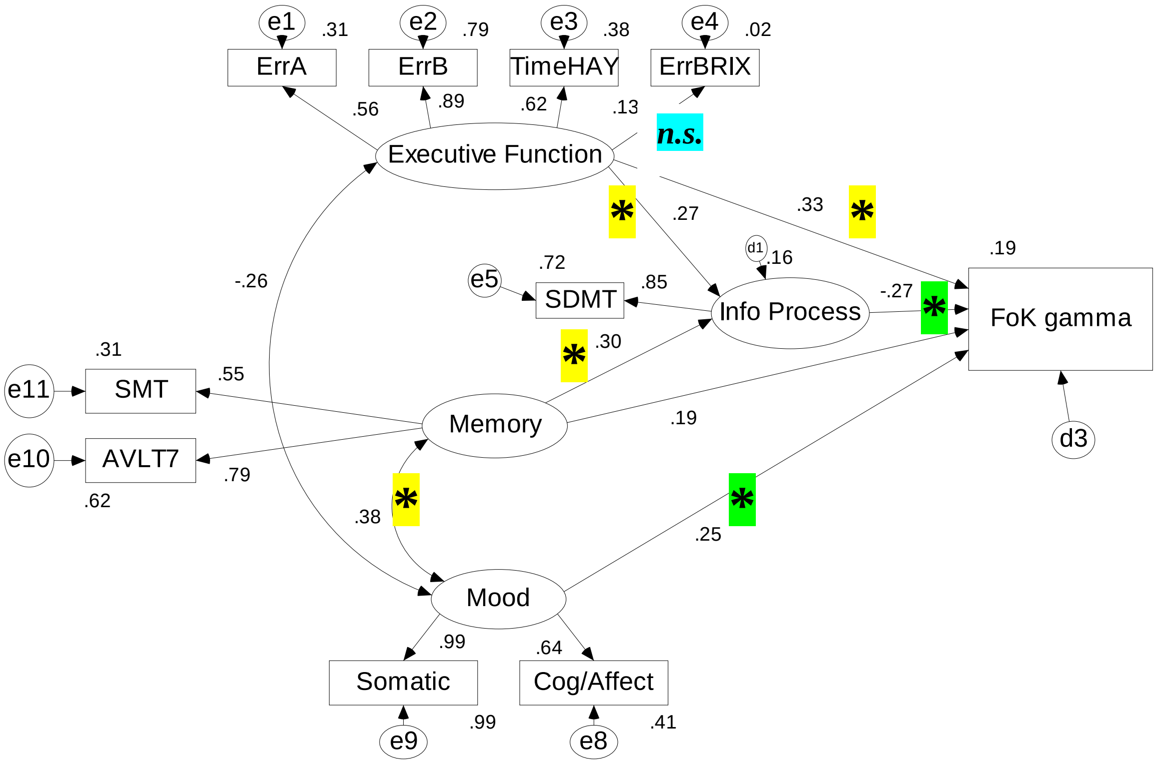


Figure 7.21 Structural mediation model of Feeling of Knowing accuracy (FoK gamma) with *Information Processing* proposed as a suppressing mediator of *Executive Function* and *Memory*. . * = significant path at p < 0.05. * = significant path at p < 0.10. AVLT = Auditory Verbal Learning Test delayed recall; SMT = Sentence Memory Tests, delayed recall; Brixton = Brixton test; Hay Time = Hayling time section B minus time Secion A; A and B errors are Hayling Sentence Completion tests errors. Cog/Affect = Cognitive Affective Beck Depression Inventory items (factot score) and Somatic = Somatic Beck Depression Inventory items (factor score). e = residuals for observed items. n.s.= non significant latent variable loading

7.3.4.10. Post hoc FoK analyses

In the FoK calibration models, two findings are of interest; the contribution of executive function in the inaccuracy condition, suggesting that poor executive function is associated with larger proportions of misapplied Low FoK ratings to subsequently recognised items. The second finding, more speculative in nature, was support for a model in which executive function may have constrained available experience-based mnemonic cues in applying FoKs, leading to higher proportions of inaccuracy. These results in part accord with the proposal of Beatty & Monson (1991) that FoK accuracy is associated with both memory and executive impairments in people with MS, and with Pannu & Kaszniak (2005), that a contribution relating to frontal lobe dysfunction appears a consistent finding in neurological populations.

In respect of the FoK relative accuracy models, a key result warrants consideration; that of a negative association between information processing and FoK gamma scores, suggesting a relationship contrary to expectations, and contrary to a non-significant zero-order correlation between FoK and performance on the Symbol Digit Modalities test. For reasons relating to the complexity of the process in deriving the FoK gamma, the gamma calculation itself, and because of the small amounts of its variance explained by the model the initial consideration for this finding is to assume the finding is artefactual. Some post hoc analyses were carried out to investigate the unexpected findings, based on a number of possibilities generated from the literature.

Heterogeneity of performance

Initially, given the distribution of FoK gamma scores, performance was split between those scoring positive and negative FoK gammas, with zero scores used to split the file. The two groups were compared (> 0.00, n=67 and < 0.00, n = 33). Comparison of mean scores across the range of observed variables was carried out. Comparing FoK subgroups, no significant differences were found between groups on memory, executive, information processing or mood variables.

A second comparison across group was carried out to reflect those who were accurate (greater than chance, n = 53), who performed at chance level (n = 26; a one in eight chance of being correct or gammas of -0.125 to + 0.125) and those who were systematically inaccurate (gamma < -0.125; n = 21). In all assessments support for the negative parameter estimate for information processing and the positive parameter value for Mood was maintained.

7.3.4.11. Summary of Feeling of Knowing models

There was general support for an executive relationship with poor performance on Feeling of Knowing judgments. This was evident in both inaccuracy in FoK calibration, and in gamma scores. While information processing might mediate the relationship of executive function and FoK gamma, the relationship warrants consideration because of the unexpected direction of the relationship between processing speed and FoK accuracy. Memory was found to be unrelated to FoK relative accuracy (gamma). Finally, mood disorder was related to accuracy in relative FoK accuracy; the stability of the relationship appeared inconsistent across the direct and mediational models however. Of note in this respect the relationships between memory and executive function were less robust, and only the inhibition-related descriptors of the latent *Executive Function* (Hayling test items) were significant contributors in the FoK gamma model.

Calibration accuracy of FoK was associated with mnemonic, and potentially inferential contributions and relative accuracy with affective disorder and executive ability, in conjunction with slower processing speeds. The relationship between memory and absolute accuracy may in part have been because of the way accuracy was measured, using only recognised sentence completions. The key result therefore was the potential contribution of executive deficit to increasing levels of inaccuracy, even in subsequently recognised items.

7. 8. General Summary.

This chapter investigated a range of metamemory measures, including memory self-efficacy, Judgment of Learning, Retrospective Confidence Judgment and Feeling of Knowing Judgments. Performance was considered specifically in relation to three factors proposed to contribute to these judgments - Memory, Executive Function and Mood. In large part results were in line with a priori models; that Mood was a contributor to memory efficacy judgments, both directly and acting in a mediation role to more inferential aspects of judgment. This confirmed the findings of Randolph Arnett & Freske’s (2004) model, using a similar instrument.

In Retrospective Confidence models, the relevance of mnemonic processes in accuracy was supported. Accuracy in Judgments of Learning at delay was related to memory ability, but not in the expected direction, with some evidence perhaps that while the learning over many trials was accounted for, the judgment might not have been. Finally, for Feeling of Knowing, a more complex interaction of factors was found; calibration accuracy was associated with memory ability, and memory and executive function appearing to conflict in their role in generating inaccurate calibration of FoK judgments. In this respect the findings of Beatty & Monson (1991) are reflected in part in these results. Mood disorder related to accuracy in relative accuracy of FoK judgments and executive contributions only disclosed when shared information processing related variance was partialed out.

Unexpected findings related to the apparent contribution of mood disorder towards accuracy in the FoK task, along with the negative association between information processing and accuracy in this task. With both results in mind, along with the contribution of executive function towards accuracy in the task, an emerging picture of dissociated contribution of information processing speed in cognitive, but not metacognitive tasks is suggested. Additionally, processes that might limit information processing - slower speed and mood disorder - appear to confer benefits in accuracy, positing a potentially important inhibitory construct in FoK accuracy.

The findings presented here have been considered only as they drive modifications to model specification, and their interpretation and relevance to the broader aims of the study will be considered in the following chapter.

Chapter 8: Discussion

8.1. Introduction.

One of the key reasons that metacognition in clinical samples should be better understood is because of its importance in self-regulated learning (Toglia & Kirk, 2000; Efklides, 2006) This learning is supported by an ability to control related processes, contingent on accurate monitoring of ongoing learning. The monitoring of memory is therefore of interest in understanding learning capacities, as it offers insights supplementary to tested memory performance. Metamemory is not typically a consideration in estimating the potential for learning, but may be more relevant to the estimation of rehabilitation potential than object-level memory (or cognitive) performance alone (Cicerone & Tupper, 1991; Cicerone et al., 2005; Ownsworth et al., 2006; Ownsworth & Clare, 2006). Additionally treating the impairment in learning rather than the memory impairment may be a way of proceeding. (Chiaravalotti et al., 2005)

Broadly, the results of this study point towards some areas of strength in memory-oriented judgment, despite varying levels of object-level memory, executive, information processing and affective status. In considering the relevance of performance across the various metamemory tasks, attention was paid to the differences between accurate and inaccurate performance on calibration measures; these absolute accuracy measures are relevant, because for the learner, they are an index of overall levels of learning (Hacker, Bol & Keener, 2008). Additionally, the contributions to relative accuracy were considered because of its importance to the learner in identifying specifically which information is known, or not (Hacker, Bol & Keener, 2008). This was considered in terms of Retrospective Confidence and Feeling of Knowing gamma scores. Taken together, relative and absolute accuracy in memory monitoring contribute to the selection and use of a range of control operations to manage performance (Nelson & Narens, 1990; Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009)

Based on the expected multifactorial nature of these judgments, the analytic approach modelled interactions of latent variables in their contribution to metamemory judgments, and each tested model was set *a priori,* informed by relevant literature. In rehabilitation, much of the reasoning carried out by clinicians might be considered to proceed at the level of latent constructs such as learning, insight, motivation and rehabilitation potential. The models investigated underscore this reasoning. While observed variables, such as performance on individual tests, may be used for diagnostic purposes, intervention and outcome is often considered and measured with respect to performance on a range of tasks required in daily life; these tasks reflect multiple domains of cognitive performance, and other, often person-controlled, variables. Latent variable modelling can potentially act as a bridge between neuropsychology and neurorehabilitation, connecting performance scores to the interactions between cognitive domains in generating behaviour. In rehabilitation these interactions are typically manipulated to provide graded, task-related, individually tailored interventions.

The overall goal of the study was to investigate metamemory in people with Multiple Sclerosis, by addressing three issues:

- First, by investigating the factors that contribute to metamemory performance in MS, both in accuracy and inaccuracy of judgments. Previous studies of the topic in this clinical group were replicated and extended using additional measures of memory monitoring and more robust statistical analysis.
- Second, the extent to which current models of metamemory account for the performance of the sample was investigated. Mnemonic experience and inferential factors associated with accuracy in metamemory were considered to this end.
- Third, the literature was extended by the inclusion of both mood and information processing as potential contributors. Specifically this related to investigating if information processing, as it does in object level memory performance in MS, functions as a constraint on performance in metamemory.

The implications of the findings for rehabilitative approaches are also addressed. At the end of the discussion, consideration is given to emerging questions and limitations of the study, and how they might be addressed. Prior to discussing the findings in detail, a summary of the study, its context, and participants is presented.

8.2. Summary of Study

The sample was generally representative of other UK community-dwelling samples of people with MS, offering some potential for generalising the results, in terms of mean age of the sample, employment status, and MS subtype (O'Hara, DeSouza & Ide, 2000; Kobelt et al., 2006). Disease subtype reflected the expected incidences. Relapsing-remitting MS participants were younger than other participants. Primary progressive MS incidence was in line with expected proportion of the sample, and the later age of diagnosis. Finally, the decreasing proportions of relapsing-remitting MS accompanied by increasing secondary progressive MS in older groups of participants is also in agreement with a changing disease course from relapsing to secondary progressive types over time (Noseworthy et al., 2000; Compston & Coles, 2002).

Cognitive impairment, including information processing, memory and executive deficits, was present in 62% of the sample, higher than some other studies, but within the range of reported incidences (Rao, 1995; Deloire et al, 2006). Findings concord with differing incidences of cognitive impairment within MS subtypes; benign MS being least, and progressive forms most, associated with cognitive impairment (Rao et al., 1991a; Rao, 2004; Denney, Sworowski & Lynch, 2005). Mood disorder was noted in about one third of participants, mainly endorsed to a mild or moderate level.

The mix of cognitive impairment across the sample reflects the common experience of heterogeneity in MS samples. However, from the individual test results presented, subgroups based on memory and executive ability did not provide clear differences in the range of metamemory measures. This finding is not entirely in opposition to Beatty & Monson's (1991) findings because in that study only one comparable measure of metamemory (FoK gamma) was reported, and groups were only reliably different in accuracy when compared to the control and cognitively unimpaired MS group, not to each other. This suggests that within impaired participants, heterogeneity of performance would in fact not differentiate subgroups of metamemory accuracy. Rather, cognitively impaired subgroups should differ from controls and cognitively unimpaired subgroups.

Contributing factors to metamemory in *a priori* models were operationalised as *Executive*, to indicate inferential and frontal processes, *Mnemonic* to indicate a range memory-experience processes and affective to indicate mood-related evaluations, which might impact judgments. Consideration was also given to information processing ability, which was expected to constrain effortful processing related to the tasks, and therefore accuracy.

Confirmatory factor analysis was used to evaluate how well selected measures could maintain discriminant validity so as to be used as latent variables. Measurement model assessment of the Beck Depression Inventory, version II supported a two-factor solution to this scale consisting of Cognitive/Affective items and Somatic items, in line with the findings of Beck, Steer & Brown (1996). This two-factor structure, the only reported confirmatory factor analysis of the BDI-II reported in the literature for this population, presented the best statistical fit; the finding concords with a number of similar investigations (Arnau et al., 2001; Viljoen et al., 2003; Thombs et al., 2008).

Of specific interest in respect of the cognitive items' measurement model was a factor model that substituted the *a priori* model. In selecting measures for the latent variables, initially a three-item information-processing factor was proposed to indicate a range of processing related features - speed, capacity and complexity. However, this was not supported, in the context of memory and executive latent variables. A better fitting model was found that supported related, but separable, *Information Processing* and *Working Memory* factors. The model in fact reflected findings of a number of studies in MS, which fail to find a coherent information-processing factor (Chiaravalotti et al., 2003; Drew, Starkey & Isler, 2009;). The theoretical orientation of the study was based on findings that information processing, and not working memory, is the primary deficit in MS. This factor, described solely by performance on the Symbol Digit Modalities Test, was therefore endorsed for assessment in the structural models.

The final set of factor analyses, and specifically that based on the General Frequency of Forgetting ten-item scale, aimed to confirm a factor structure for the memory efficacy scales. Generally, there appeared scope to reduce these scales further because both appear to have redundant items. There is also a possibility, in the use of Forgetting while Reading items, that factoral coherence related to common method effects, rather than to their strong association with efficacy. Perhaps one of the benefits of both scales is that they do ask for an assessment of specific memory failures; the General Frequency of Forgetting scale relating to a more representative range of memory failure instances. While not high, missing data instances on these scales, and the Memory Function Questionnaire more generally seems to reflect some items that are inappropriate to some participants. Items asking about memory in public speaking, test-taking and reading generated missingness, suggesting reduced applicability in the population. The results from both factor analysis and structural modelling supported the use of the ten-item General Frequency of Forgetting scale in favour of the alternative.

Confirmed latent variables were then tested in a set of *a priori* structural models to examine the relationships between these and a range of metamemory judgments; memory self-efficacy, Judgment of Learning, Retrospective Confidence Judgment, and Feeling of Knowing. With clinical application in mind, the study's focus was on episodic memory, in contrast to a sizable amount of metamemory research, which has typically focused on semantic memory.

8.3 Metamemory in Multiple Sclerosis

Given that there are few established relationships between different measures of metamemory (Souchay, 2007; Dunlosky & Bjork, 2008a; Dunlosky & Metcalfe, 2009), each is considered initially as a separate judgment, prior to considering them as a set of conceptually related monitoring judgments. This discussion generally relates to the results of structural models, and will focus on the questions for the study. These regarded contributory factors to each judgment, how results situate with current models of each judgment and the role of information processing and mood in metamemory accuracy.

8.3.1. Memory Self-report

The first finding from this study was that subjective memory report (memory self-efficacy) was not found to relate to tested memory performance. Instead, mood was a reliable independent predictor of memory complaint, supporting the findings of a number of studies of both neurologically impaired and neurologically intact samples (Broadbent et al., 1982; Lovera, et al, 2006).

On the General Frequency of Forgetting scale, participants reported more problems, and considered these more serious when they happened, than non-neurologically impaired people, even though the normative samples, against which they were compared, were older. An older sample might typically expected to endorse more memory difficulty than younger samples, in part because of implicitly held theories about aging and memory decline (Hertzog, 2002). In addition, all cognitive subgroups (including the no memory impairment group) rated themselves as having more memory difficulty than age-matched groups from the complete Memory Function Questionnaire standardisation sample. Alone, mood disorder established a reliable association with memory self-efficacy, with higher depressive symptomatology being associated with more memory complaint. There was some indication too that evaluative aspects of efficacy judgments may have been mediated by this mood bias.

In concluding that this index of metamemory might be impaired, consideration should be given to the fact that objective memory impairment was evident in the sample. Defining accuracy in memory efficacy would suggest it should relate to tested memory ability. This perhaps ignores both the impact of extraneous contributors to efficacy judgments, and the experiences of memory failure. Two anecdotal findings both from study participants are relevant here. Word-finding difficulties and other experiences of apparent forgetting, are to the person experiencing them, often considered as evidence of memory problems. In reality they might better be considered non-memory deficits. Semantic and phonemic fluency problems in MS are common and appear to relate to processing speed and executive abilities, rather than memory impairment (Henry & Beatty, 2006), despite being experienced as such. Likewise, apparent forgetting, may relate to acquisition failures, again associated with processing resources in MS (Lezak, Howieson & Loring, 2004).

Of interest therefore, was a failure to find a relationship between processing abilities and memory self-report when this was tested in the structural models, suggesting a complex and indirect relationship in how experiences attributed to memory failure may be incorporated into memory-efficacy, or stored metacognitive knowledge of memory. One such possible mechanism might relate to mood disorder; each experienced memory problem confirming already negative memory evaluations about memory.

In assessing the two memory-efficacy models, there was support for the findings of Randolph, Arnett & Freske (2004), that mood acts as a mediator of the evaluative components in making this efficacy judgment. Those with higher levels of mood disorder reported more memory problems in daily life, and the contribution of executive function was reliably different when examined as a mediated contributor, compared to a direct effect. The question remains of how that relationship might be understood.

One theme of studies, from samples of older people with late life depression (Alexopoulos et al., 2000), is that the relationship between both mood and executive function may relate to disruption of a number of cortical-subcortical loops, contributing to an imbalance of excitatory and inhibitory modulation of both cognitive and affective function. In MS, left hemisphere pathology, specifically left frontal and temporal lobe areas appear implicated in depression (Feinstein et al., 2004; Siegert & Abernethy, 2005). These findings offer potential for a neural correlation between executive and affect disturbance. Mechanisms by which depression and executive function might interact include difficulties in shifting cognitive or emotional set of focus (Bunce, Handley & Gaines, 2008; Wang et al., 2008). This might relate to ‘state’ components of depression such as the impact of rumination on inhibitory control (Watkins & Brown, 2002). Negative biases in thinking more generally, which might relate to ‘trait’ views of the impact of depression (Beevers & Miller, 2004). Each proposes an extent of mediation by mood of evaluative processes.

Experimental evidence for such mediation is proposed from the treatment of mood disorder leading to changed perceptions of cognitive ability in MS (Julian, Merluzzi & Mohr, 2007). Mechanisms for such findings could relate to changing the cognitive bias caused by mood disorder, or attenuation of the impact on mood on effortful cognitive processing, noted in MS (Arnett et al., 1999b; Demaree, Gaudino & DeLuca, 2003). In the study of Julian, Merluzzi & Mohr (2007), the authors propose a relationship between objective cognitive function and subjective report emerging as a result of the attenuation of mood disorder. This suggests a causal role of depression in the relationship between objective and subjective measures. In people who are cognitively unimpaired, mood disorder has also been suggested to be associated with complaint of cognitive difficulty (Broadbent et al., 1982).

A second area for consideration in respect of these results is whether there are other specific explanations for the lack of a robust relationship between memory self-report and tested memory performance. Questioning about memory ability is often the initial assessment of memory ability made by clinicians. Memory complaint is often used as a trigger for formal memory evaluation, so such a relationship is often assumed. Despite the structural models’ findings of no reliable relationship between tested and reported memory performance, an additional finding was that memory efficacy scores did correlate with performance on one measure of memory used in the study, but not the other. The two memory tests may reflect different aspects of experienced memory failures - the single trial

Sentence Memory Test, acquisition and the multi-trial Auditory Verbal learning Test, retention.

While the experience of people with MS may be of memory failures, there is some evidence, with which Sentence Memory Test performance and efficacy correlation accords, that difficulty with acquisition, rather than retention is a key memory problem in MS (Lezak, Howieson & Loring, 2004). The correlation, albeit low (Pearson's r = 0.24), for each memory efficacy scale and Sentence Memory Test, suggests some sensitivity in the questionnaire to actual memory performance. The developers of the 10-item General Frequency of Forgetting (GFF) scale have proposed that memory performance explains small amounts of the variance in GFF scores, even with individual factors controlled (Zelinski & Gilewski, 2004). In the study of Randolph, Arnett & Freske (2004) no relationship between memory-efficacy and tested memory performance was found, but the memory measure used was one in which participants were trained to a criterion level of performance before memory was tested, thereby negating the impact of acquisition problems. From the results presented in this study, a focus on acquisition deficits in memory testing may yield a more consistent relationship between efficacy and tested ability in people with MS.

The lack of a consistent association between self-reported and tested memory performance might also relate to factors, aside from the type of memory process tested. Experienced memory failures may be infrequent because environmental demands have been adapted downwards to equate with disability; on standard assessment of memory impairments might be sizable, but complaint low (Rabbitt et al., 1995; Hertzog, 2002). The reverse proposition might better reflect the finding in this study, given memory complaint was higher than normed samples. This could imply a failure to adapt environmental demands in the context of experiences of memory deficit, this perhaps suggesting inadequate control processes.

Tests of memory may not be ecologically valid in comparison to a questionnaire, or there may be insufficient specificity to actual memory behaviour in both questionnaires and memory tasks (Hertzog, 2002). This may point to ways in which questions about memory in MS might be improved by subjective memory report specifying a more task-specific 'retrieval context' (Hertzog, 2002:184) to limit the use of generalised self-appraisal.

Hertzog, for example, suggests that questionnaires relating to specific memory behaviours such as remembering to take medication should be related to measured instances of ‘forgetting in medication taking’ to disclose the self-report and tested memory link.

Additional factors attenuating the self-report and tested memory link in MS samples might also include factors that apply in non-clinical groups too - such as age (Hertzog, 2002; Lineweaver & Hertzog, 1998) or personality factors (Zelinski & Gilewski, 2004). Inclusion of these factors in an assessment of the contributors to efficacy judgments might also explain increased amounts of variance in efficacy; here, 23% of the variance of the preferred General Frequency of Forgetting scale was explained, suggesting inclusion of additional fators is warranted.

Memory efficacy might best be considered in respect of how efficacy judgments can be calibrated, in order that they do not limit memory effort, rather than approached from an assumption that accuracy is quantifiable. It is perhaps because of these validity issues that memory-efficacy measures are not the focus of the majority of metamemory research. Importantly however, efficacy is seen generally to have an important role in the sense of agency in human activities, meaning memory efficacy probably influences learning-oriented effort, regardless of correspondence with objective measures (Bandura, 1982; Sniehotta, Scholz & Schwarzer, 2005). Rather than self-report being indicative of memory performance, it may have validity in predicting memory or learning effort.

8.3.2. Delayed Judgment of Learning

Accurate delayed Judgments of Learning are important in decisions about how much is known (Metcalfe & Finn, 2008). The results here indicate underconfidence with practice was maintained at delay, that better recall was contributory to greatest underconfidence and that information processing speed was associated with greater accuracy. The proposal in assessing this contribution is that both may reflect those with better *object-level* cognitive function.

The accuracy measure used in this study was the discrepancy between predicted and actual recall, after a delay, and all scores were negative, indicating underconfidence in the judgments. The presence of both initial overconfidence and underconfidence-with-practice across learning trials and are common findings during list-learning among unimpaired samples (Koriat, Sheffer & Ma'ayan, 2002). In this regard the MS sample displayed typical patterns of learning during the task.

One model of how accuracy in JoLs after delay in achieved, compared to pre delay, is the Monitoring Dual Memory (MDM) hypothesis (Nelson & Dunlosky, 1991; Koriat, 1997; Narens, Nelson & Scheck, 2008; Dunlosky & Metcalfe, 2009). The MDM model proposes that an immediate JoL is based on monitoring short and long-term memory availability and delayed JoLs are made only from long-term memory monitoring. Delayed JoL is proposed to be a more accurate indicator of learning, because covert retrieval attempts are made from the same memory store used for the recall test, increasing concordance between prediction and performance. The main implication in assessing the JoL results relates to the presumed use of covert retrieval attempts to abolish underconfidence with practice at delay (Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009). This was not the finding here however.

A key finding was of maintained underconfidence in the delayed Judgment of Learning (JoL). This phenomenon not typically reported in the literature. A second finding suggests that most underconfidence in the judgment was associated with better recall performance. A number of possibilities for explaining these results come from two perspectives of how this judgment might be implemented in unimpaired samples. Broadly, these relate to failure in using the correct diagnostic cues and artefactual mnemonic factors, relating to the judgment task itself.

It was seen in the structural model that prior JoLs were contributory to the delayed JoL, supporting a contention that pre-delay underconfidence contributed to the error of low recall prediction after a delay. Nelson & Meeter, (2003) suggest that *'perhaps when making JOLs people do not give as much weight to external cues (such as the amount of study time) as objectively should have been given'* (2003:131). Here, the failure may have been not attending to the likely benefits of five learning trials, but instead indexing the delayed judgment against mounting underconfidence from earlier trials. Meeter & Nelson, (2003) and Finn & Metcalfe, (2007) suggest a reason for maintained underconfidence might be mistaken assumption that immediate judgments of learning are better indicators, not only of learning, but also of subsequent Judgments of Learning. The structural model does support a reliable relationship between pre-delay and delayed JoL. However, in proposing this explanation of the findings, only the delayed underconfidence itself is addressed, not the finding that larger discrepancy (more underconfidence) was associated with better recall performance.

An alternative theory about delayed JoLs might offer some explanation of that result; the so-called *self-fulfilling hypothesis* (Spellman & Bjork, 1992). According to this model of delayed JoLs, if covert retrieval is carried out in making JoLs there is better learning from which a final covert retrieval can be made, abolishing the underconfidence effect. The retrieval practice provided through multiple trials, may lead to a large increase in final recall (Narens, Nelson & Scheck, 2008), explaining the increase in performance in better rememberers. The wider discrepancy between prediction and performance, begs the question as to why prediction did not increase in line with better performance, if covert retrieval took place. The proposal in interpreting the results presented here is that the influence of prior JoL predictions accounts for the maintained underconfidence, which was accompanied, in better rememberers, with a *self-fulfilling* increase in learning, so widening discrepancies.

Finn & Metcalfe, (2007; 2008) suggest a *memory for past test* heuristic as a cause of underconfidence with practice effects because people use prior performance as an indicator of subsequent performance, and it may be that this was extended to delay. With the inclusion of the additional relationships in the final structural model, immediate JoLs contributed to underconfidence in delayed JoL, memory ability was positively associated with accuracy pre-delay JoL, but negatively in delayed JoL. The picture that emerges is of delayed JoL accuracy being associated with the JoL prior to delay (‘what did I get last time’), itself containing underconfidence. If this underconfidence served to lower predictions, then those who were in fact learning better would generate wider discrepancies between predicted and actual performance. This would give the appearance of more underconfidence in those with better memory, but the discrepancy in facts relates to better recall performance relative to others. Nominally ‘less accurate’ delayed JoL is in part therefore, artefactual because the multiple learning trials offer spaced retrieval practice, from which better rememberers benefit most (Kimball & Metcalfe, 2003).

Perhaps then, the more interesting finding was the maintenance of underconfidence, regardless of its size. The relevance of maintained underconfidence is that it might indicate poor task apprehension or attention to non-diagnostic cues. Use of cues including what might be considered a memory-efficacy effect (prior JoL), may have driven the judgment more than appropriate mnemonic control operations, such as covert retrieval. The efficacy contributions might be understood as a conflict between what Vilkki, Servo & Surma-aho, (1998) term *stored metacognitions* and *concurrent metacognitions* about memory - the ideas people have about their memory, or a task, compared to their within-task monitoring of memory. Participants may have taken a global approach to making the judgment, which may have included such extrinsic, information-based factors; prior performance or heuristics about the impact of delay on memory, rather than using available mnemonic methods to make this judgment (Meeter & Nelson, 2003). In line with Koriat's (1997) proposal of the cues relevant to JoLs, specific attention may be required in weighing up intrinsic, extrinsic and mnemonic cues, and an understanding that some cues might assist with memory, and others with a Judgment of Learning. This final point may imply a failure apprehend the task context fully, and implicate some wider executive failure. This may relate to the proposal of Van Overschelde, (2008) that monitoring processes should be considered in terms of their goal basis, rather than a passive process.

It is not known from this study, whether covert retrieval was used in making this judgment. Evidence from the high Retrospective Confidence Judgment accuracy of the sample suggests that it would be an accurate predictor of performance, so it appears for the delayed JoL this may not have been a failure of ability to accurately monitor what is known. Instead it implicates failure to do so, because of inattention to the key requirements in making this judgment.

One final consideration is of the role of information processing in respect of accuracy in the delayed JoL. Bearing in mind that the judgment remains debated in terms of whether it is fundamentally mnemonic (object-level) in its nature (Kimball & Metcalfe, 2003), the findings that processing speed is associated with accuracy might support a relationship with object-level performance, the improved list-learning over trials. This would be akin to the proposed role of information processing mediating other object level cognitive processes in MS. Information processing maintained a positive and reliable correlation with the other cognitive domains as would be expected, but while speed was associated with greater accuracy in the delayed JoL, better memory performance was associated with less accuracy. One potential explanation is in recourse to the object-level contribution the judgment. Both memory ability and information processing speed were associated with better performance, but only in relation to the performance components of the Judgment, not the predictive component. This interpretation, though speculative accounts for the positive correlation between the two variables, while also accounting for their apparently opposing relationship with accuracy at delay.

In summary, the maintained underconfidence at delay on the JoL task was a relatively novel finding, supporting the maintenance of pre-delay JoL judgment in informing the delayed judgment. Accuracy was associated with memory ability in a direction opposite to that expected in the a priori model. Explanations for this might relate to increased learning over study trials being greater than appreciated by participants, appearing to suggest some impairment in monitoring. The result might also indicate, in agreement with Kimball & Metcalfe (2003), that the process of making the JoL may not be a clear metamemory judgment.

Attention to the diagnostic task characteristics may have been a difficulty, and this may relate to poor task appraisal or to the structure of the task itself, appearing to emphasise learning, rather than judgment about learning, as its goal. Whether the interpretation relates to the task design, or to participants failures to use diagnostics relevant to accuracy appraisal, the importance of apprehending the key information from a task is underlined. The importance of this task appraisal might therefore be characterised as paying sufficient attention to intention (Lau, et al., 2004).

Proposals that delayed Judgment of Learning is primarily a memory event in neurologically unimpaired samples perhaps assumes that the initial task-apprehension stage is not part of its metacognitive status; here it appears that the failure in this regard could be attributed to some broader executive ability. While this is not supported by a statistical relationship between delayed Judgment of Learning and executive function in the tested model, the results may suggest a failure to deploy monitoring appropriately towards key diagnostics for accuracy.

8.3.3. Retrospective Confidence

Participants were accurate at making judgments about which items were more likely correct than others, after recall attempt, even on a single exposure new-learning task. The

Retrospective Confidence Judgment (RCJ) task has traditionally established high levels of accuracy even in clinical populations, such as people with traumatic brain injury (mean RCJ gamma = 0.99; Kennedy 2001). It is possible that the high levels of relative accuracy in RCJs relate to the use of an episodic task because there is limited availability of alternate information from which to make a judgment about confidence. This is contrary to a semantic version of the task, which is likely to require the assessment of a number of candidate answers, because of its general knowledge basis (Perfect & Hollins, 1996). Given a general knowledge question for example, it is likely that there would be a number of alternative candidate answers that might confound making a confidence judgment, in conjunction with a general assessment of domain expertise. High levels of accuracy in the confidence judgment found in this study, suggest that this process could be considered a reliable source of information in new-learning (episodic) situations, so that self-testing approaches might provide a useful index of what is known (Kennedy, 2001).

In terms of absolute accuracy, memory ability was a contributor to the proportion of correctly recalled targets given a high confidence rating. For the few instances where correctly retrieved targets were given a low confidence rating, none of the selected latent variables were explanatory. Failure to find relationships therefore may have related to their small number across the sample or the lack of a systematic reason for their occurrence.

Despite the high levels of accuracy, one potential limitation in recommending the post-retrieval confidence judgment as a reliable measure of knowledge in episodic memory relates to the findings of Brewer & Sampaio, (2006) and Brewer, Sampaio & Barlow, (2005). Results from these studies suggested that this confidence judgment was prone to specific types of errors. Their experiments demonstrated that the inclusion of incorrect, but gist-related, sentences into a judgment process caused over-confidence. They proposed that if there was mistaken recall of having seen the deceptive sentence (of similar gist) during the original study episode, confidence was higher than warranted. This biasing of confidence in retrieved material may also extend to how responses are sought during tasks. Forced report, where people are not allowed a don’t know option, means that effort is often required to generate answers, and this effort, plus the memory experience of the given answer itself, may act to raise confidence in this answer in a future situation, regardless of its accuracy (Koriat & Goldsmith, 1996).

A positive interpretation of this relationship between effort and confidence judgment is that in a sufficiently structured, or even errorless, forced report condition, self-generation may assist learning (Akhtar, Moulin & Bowie, 2006). One MS related study that used self-generation to guide recall of the steps of a task, reported improved performance (Goverover, Chiaravalotti & DeLuca, 2008). The intervention aimed to teach the sequence of everyday tasks by providing incomplete sentences, which supported generation of the correct steps of the task in a highly structured way; an example for an omelette preparation task was 'beat together two___'. In terms of design, such a training approach would fulfill the criteria outlined above for a structured, errorless, forced report intervention. While the study in question proposed the method as enhancing encoding, it could additionally be seen as one in which effort at encoding would increase fluency experiences at recall, indicating confidence, which is predictive of performance in recall.

In free report paradigms, as in this study, where *don't know* answering is allowed the limitations of self-generation may be avoided (Dunlosky & Metcalfe, 2009). In rehabilitation contexts, a *don't know* option is an important component of learning monitoring, and given its utility in increasing accuracy, it is appropriate. One limitation in regard to recommending free report as a method for avoiding confidence in non-studied information, is that it presupposes that people can themselves place the report/don't report criterion at the appropriate level by themselves, that is have good monitoring and control processes. Executive dysfunction in this sample was evident in about one third of participants and such dysfunction has been associated with difficulties in controlling retrieval (Thornton & DeFreitas, 2009). This may predict that the placing of the report threshold could be difficult; an extreme example being confabulated memory responses (Johnson et al., 2000). Despite this, on item-by-item recall, this did not present cost in terms of accuracy. This relative accuracy finding is mirrored when calibration is considered too, because there were low instances of inaccuracy.

In both assessment of relative and absolute accuracy, retrospective confidence was associated solely with memory performance. The result supports the view that privileged access and mnemonic cues support accuracy in these judgments, and that some inaccurate judgments (that is low confidence even when recall is correct) are unrelated to memory performance. This may suggest that these are instances where although the target is available, and is recalled, the ascribing of a low confidence might be based on insufficient retrieval fluency or memory-experience to generate a confidence assessment. This interpretation might concord with the wider variability in accuracy scores observed in subgroups with memory impairment.

The positive relationship between memory performance and accuracy is consistent with the proposal that those with better memory performance have more robust memory-experience cues (Wheeler, Stuss & Tulving, 1997; Wheeler, 2000; Brewer, Sampaio & Barlow, 2005; Koriat et al., 2008). For relative accuracy, cognitive subgroups were no different suggesting that memory performance *per se* was not a significant factor, though those with memory impairment alone tended to be more variable in their performance, perhaps supporting a characterisation of their memory-experiences as being of less diagnostic utility. In terms of absolute accuracy, some support was available to make an interpretation that memory impaired subgroups (Memory impaired, Memory & Executive impaired) may have reduced confidence for correctly retrieved targets, again supporting a mnemonic-experience basis for this judgment. This is consistent with the findings of Maki (2008) that privileged access accounts for much of the judgment material for Retrospective Confidence Judgments, specifically ease and fluency in retrieval experiences (Benjamin & Bjork, 1996; Koriat et al., 2008).

Findings about Retrospective Confidence in this study demonstrate that post-retrieval confidence to be a reliable indicator of what is known, in episodic memory, under conditions of free report. The sample, when given a single learning trial could accurately identify the differences between high, medium and low levels of confidence, and this related to actual retrieval success. This offers support for a view that this metamemory judgment is accurate, even in people with impairments in memory, executive function and mood disorder. The judgment is likely to have benefited from a number of factors - the time they were made (after retrieval), the item-by-item structuring of the task and its episodic nature. Only memory-related processes were reliably associated with accuracy, underlining the presumed importance fluency and availability likely being two important cues to such accurate judgments. No reliable relationship was found between accuracy on this judgment and information processing abilities, which might support the view of Koriat et al., (2008) that the process is parasitic on object-level memory processes and therefore does not require additional controlled processing.

8.3.4. Feeling of Knowing - Relative accuracy.

As a group, the finding of a mean FoK gamma of 0.18 suggests a low level of accuracy. Some participants did reach gamma correlations of 1.0, and the modal score was 0.50. Two studies, which report a FoK gamma correlation on a similar task, are those of Shimamura & Squire (1986) and Schnyer et al., (2004). Shimamura and Squire compared a control sample to a sample with a range of memory impaired impairments; amnesics, people receiving ECT and people with Korsakoff's syndrome, the latter associated with additional executive deficits. FoK performance showed controls achieved a mean FoK gamma of 0.70 and amnesics 0.40, both higher than this study's sample. The Korsakoff's group achieved a mean FoK gamma of zero. The authors concluded that more widespread cognitive impairment is probably therefore responsible, rather than relatively circumscribed memory impairment, for success in this task. This accords with the findings in this study that the lowest levels of accuracy, both in terms of relative accuracy (gamma) and absolute measures focusing on inaccuracy, being associated with executive dysfunction.

As a group therefore, this sample had low gamma accuracy values in the FoK task, supporting a view that this is impaired. In the sample of people with MS reported by Beatty & Monson (1991) mean FoK gamma is reported for each cognitively impaired subgroup and ranged from 0.24 (SD = 0.32) down to -0.02 (SD = 0.35) across the sample. Given the findings of Beatty & Monson that lowest gammas were associated both with memory and executive impairment, the results here offer only partial agreement. Only executive impairment, and not memory, was reliably associated with relative accuracy here. This supports an interpretation of some dissociation of memory and metamemory ability (Souchay, Isingrini & Gil, 2006; Souchay, Bacon & Danion, 2006). Considerations in respect of differences in approaches to calculating gamma may be relevant, and will be discussed later.

Schnyer et al., (2004) report a sample of people with frontal lobe damage, but do not detail the mean FoK gamma statistic for the group as a whole. Instead only highest (mean gamma = 0.41), and lowest performers (mean gamma = - 0.31) are reported. Their findings also support a relationship between relative accuracy and executive ability, but not memory.

The methods for acquiring FoK judgments were very similar across Beatty & Monson (1991), Schnyer et al., (2004) and this study, so differences in the calculation of gamma may be a factor in the different findings. Beatty & Monson (1991) do not explicitly state the method of calculation of the gamma score, though the suggestion from the paper is that it was based on the *sum* of concordances minus *sum* of disconcordant FoK as a proportion of all judgments. Some have argued that this product-based calculation of accuracy in fact produces the Hamann coefficient, a measure of absolute accuracy (Nelson et al., 1986; Nelson, 1996). A comparison with the absolute accuracy results therefore may be more appropriate, in which case there is support for the memory and additional executive impairment being associated with the lowest accuracy condition here (higher numbers of Low FoKs correct), as was suggested by Beatty & Monson (1991). What the Hamann statistic measures, that is different to the Goodman & Kruskal gamma statistic, remains debated in the literature (Goodman & Kruskal 1954; Nelson 1984; Schraw, 1995; Nietfeld, Enders & Schraw 2006).

The literature points to FoKs, being based on cue familiarity and in this sense having a more inferential quality (Koriat, 1993, 1994; Dunlosky & Metcalfe, 2009; Leonesio, 2008). Inferences about future recognition may in fact be based on a range of cues, including memory of the encoding episode itself, any extant feelings of familiarity or accessibility, task features, perceived ability heuristics and perhaps cues generated by the person themselves (Reder & Ritter, 1992; Metcalfe, Schwartz & Joaquin, 1993; Reder & Schunn, 1996; Benjamin & Bjork, 1996; Souchay, 2007; Maki, 2008). These cues are evaluated for diagnostic utility in predicting future recognition in the absence of direct-access as a mechanism for making the judgment.

There were a number of relationships of particular interest in the relative accuracy of FoK models. First, the contribution of executive, but not memory ability. Second, the association of faster information processing with lower FoK accuracy, in the context of a positive relationship between executive function and accuracy. A third relationship of interest was that between lower mood state and degree of accuracy. The structural model suggests that executive ability is positively, and information processing, contrary to the assumed effect, is negatively associated with accuracy. Finally, lower mood also appears to contribute to the accuracy of FoK judgments.

These may be important indications of how Feeling of Knowing judgments differ from other metamemory judgments (e.g. RCJ, JoL), the latter proposed, and validated to some extent in the study, to have their basis in mnemonic experiences. The role of executive function concords with documented relationships in both MS and other neurological conditions (Shimamura & Squire, 1986; Beatty & Monson, 1991; Souchay et al., 2003; Pannu & Kaszniak, 2005). The contribution of information processing and mood have not been reported in models of the Feeling of Knowing in MS, so these have limited provenance to guide interpretation. A wider consideration of executive function, beyond representing inferential abilities, to indicating frontal integrity more generally, may help in understanding how the mood and information processing findings could be understood.

Frontal lobe dysfunction has been shown to reduce the accuracy of memory predictions (Janowsky, Shimamura & Squire, 1989; Vilkki, Servo & Surma-aho, 1998; Shimamura, 2000; Souchay et al., 2003; Pannu & Kaszniak, 2005; Schwartz & Bacon, 2008), and when tasks are most difficult the requirements of frontal-mediated monitoring are perhaps most critical (Pannu & Kaszniak, 2005). While there is some evidence for an association between FoK accuracy and memory ability, this has been shown to be negated when executive contributions are statistically are removed (Schnyer et al., 2004). Here the structural models statistically separated their presumed shared relationship with information processing ability, by modelling that relationship through the mediator.

In an assessment of the neural correlates of the Feeling of Knowing judgment, not only are frontal contributions generally supported, but a shift from stimulus-oriented to internally-oriented cognition is also suggested, in conjunction with a shift away from stimulus-driven attention (Chua, Schacter & Sperling, 2009). Frith & Frith, (1999) suggest that internally oriented reasoning involves ventromedial regions of the frontal lobe, an area also associated with Feeling of Knowing performance (Schnyer et al, 2004; Modirrousta & Fellows, 2008). An additional dependence of Feeling of Knowing performance might relate to a frontal role in ‘dynamic filtering’ (Shimamura, 1996). This filtering functions to control information processing, through the management of the amount and type of processing carried out. The relevant deficit then might be a regulatory one (Benedict et al., 2001).

One example of this filtering provided by Shimamura (1996:155) is the effect of interference during retrieval causing memory-blocking effects leading to tip of the tongue experiences and poor verbal fluency. As mentioned before, such experiences are common in MS (Henry & Beatty, 2006) and are often complained of as memory deficits. Frontal dysfunction, such as that indexed in this study, has been associated with a failure to inhibit processing unrelated to task demands. Such unconstrained processing is one approach to describing MS-associated disinhibition in the literature, where it is considered as evidence of a neuropsychological basis for disinhibition, in respect of fronto-limbic dysfunction (Benedict et al, 2001; Fishman et al., 2004).

The breakdown in applying control to behavioural output, including cognitive behaviours, may mean that processing abilities are poorly directed in respect of task demands. This failure to fashion responses in accordance with the task context was indicated as potential basis for maintaining underconfidence with practice in the JoL findings from the study. Additionally, it may offer a potential way to consider how speeded information processing ability may not necessarily support the internally-oriented processing required in the FoK task, where the traces and cues upon which accuracy might be based are scant. Additional processing may act as interference, as it does in controlling word retrieval in MS (Henry & Beatty, 2006). With this interpretation, the suppression effect of information processing speed on executive function, found in the Feeling of Knowing model may be considered one where additional information processing acts as interference (in content and quantity). Potentially this might lead to pitting familiarity sensations against explicit cues to future prediction, the latter hindering this kind of judgment, because of its implicit nature (Reder & Schunn, 1996).

Notably, one observed measure for the executive function latent variable became non-significant in this model, leaving the latent to be described only by the three items reflecting inhibitory processes, measured by the Hayling Sentence Completion task. In people with MS, comparing their report of cognitive ability to an informer whom they knew, it is proposed that inaccurate estimation of abilities - here indicated by the lowest FoK gamma scorers - was associated with less depression, more cognitive impairment, disinhibition and euphoria (Benedict et al, 2001, 2004; Carone et al., 2005). This could be considered a reasonable characterisation of the final structural model. Executive impairment co-occurs with fast, but unconstrained, information processing, and less depression symptoms.

Such a characterisation may reflect the features of ‘euphoria sclerotica’ (Cottrell & Wilson, 1926:8) a noted presentation of MS involving mood elevation, disinhibition and more significant cognitive impairment (Cottrell & Wilson, 1926; Kendall et al., 1987; Finger, 1998; Fishman et al., 2004). Returning to the idea of an affective continuum discussed in Chapter 3, the lowest depression symptom scorers in this context might be considered those who are *absolutely not-depressed,* as opposed to being euthymic in their mood (Kendall et al., 1987). This has been discussed as ‘low-end specificity’ in the tool (Joiner, Schmidt & Metalsky, 1994).

The interpretation of executive and information processing relationship to accuracy in the FoK task is not to suggest that there is some kind of speed-accuracy trade-off. Instead, in a complex inferential metamemory judgment with limited cues, *effective* or goal-relevant processing can be conceived of as resistance to interference. This was not the proposal in the a priori model, based on the relationships between information processing and object level cognitive functions in MS. The results from modelling suggest that in considering this meta-level process, information processing contributions may be qualitatively different to their relationship with object-level processes, and this may be because of failures in the top-down (executive) directing of processing resources.

Parallel debates as to the key deficit underlying memory decline and metamemory in older people offers some related propositions (Salthouse 1996; Bunce & Macready, 2005), with support for the processing speed decrement rather than for competing views such as executive impairment models (Balota, Dolan & Duchek, 2000). In one study, offering conceptual support to the findings here, a double dissociation was proposed between age-related variance in Feeling of Knowing accuracy and cued-recall performance. Variance in FoK accuracy was best explained by executive capacities, and in recall by information processing speed (Perrotin et al., 2006).

One way in which this might be considered is in terms of whether explicit or implicit cues are used for this task. Reder and others, in a series of papers focusing on the relationships between metamemory and implicit memory, propose that the two functions may have more in common than might first be acknowledged in the literature (Reder & Schunn, 1996). The relevant cues might be implicit rather than explicit, especially given its episodic nature. Interference in, and caused by, deliberation, itself making explicit what might be an implicit process might also offer some explanation why, even in unimpaired samples, Feeling of Knowing accuracy is typically low. This is a monitoring judgment that is implicit in nature and bringing to bear substantial controlled processing may obscure the judgment-maker from its more procedural basis. In this sense, the proposal at the outset of the study that executive function represented explicit, controlled and deliberative processing might be expanded to include the maintenance of freedom from interference so that the appropriate experiential sensations from ongoing, and perhaps implicit, cognition can be drawn into FoK judgment making. While additional information may become available to the judgment maker from this deliberation, the question of its utility has to be considered.

Informational deficiencies have been proposed as a way of understanding accuracy errors in social psychology also and may offer some conceptual support. This view of accuracy errors considers informational availability not just in terms of insufficiency, but also surplus (Kruglanski, 1989:400). An extrapolation might suggest that extensive and fast information processing might indicate to a judgment maker a level of knowing, leading to reduced vigilance towards inconsistent information (Cantor & Kihlstrom, 1987).

A final consideration is in respect of mood disorder, which demonstrated a negative association with Feeling of Knowing accuracy. The majority of participants were of a minimal/none or mild level on the Beck Depression Inventory. In addition to the suggestion proposed earlier that this could include some participants with euphoric presentations at one end of an affective continuum. People with MS who have milder cognitive impairment may be those with higher incidences of depression (Landro, Celius & Sletvold, 2003; Siegert & Abernathy, 2005). This is necessarily a tentative proposition because of the evidence of interactions between mood disorder and processing abilities, planning abilities and working memory in other studies (Arnett et al., 1999b; Arnett, Higginson & Randolph, 2001; Arnett et al., 2002). Relationships in the model could therefore also represent Feeling of Knowing accuracy, as a marker for milder or absent cognitive impairment, being associated with more depression. Further, there is evidence of dysphoric mood being associated with decreased recollective experience, but not familiarity experience (Hertel & Milan, 1994). This might support a view that, in the context of better executive performance, the bias towards familiarity processing might support accuracy, because diagnostic evaluation of familiarity experiences is both appropriate and possible.

8.3.5. Feeling of Knowing - Absolute accuracy

As with Retrospective Confidence, an attempt was made to investigate the factors that are associated with separate measures of accuracy and inaccuracy for the Feeling of Knowing judgment. This means that concordant judgments (High Feeling of Knowing, and recognised) and disconcordant (Low feeling of Knowing, and recognised) were modeled. From the results of this modelling, memory performance related positively to proportions correct of both High Feeling of Knowing and Low Feeling of Knowing. This might be expected since both measures are based on successful recognition performance.

For sentence completions correctly recognised, with a High Feeling of Knowing, this suggests that memory experience does support FoK accuracy on an item-by-item assessment of likely recognition, which may relate in part to the use of cue-familiarity assessment. For inaccurate assessments of future recognition, a model in which executive function mediated memory performance to negative effect was assessed. Although having no statistical superiority to a direct-effects model in characterising the data, it may clarify the relationships between the two. That the key deficit in leading to greater inaccuracy is the frontal mediated limitation in the assessment of future recognition.

The absolute accuracy mediation model proposes that, despite subsequent recognition performance, the application of a Low FoK rating may be associated with executive impairment; a failure to usefully evaluate cues diagnostic of future recognition. The mediational model will require validation as it accounted for the data only to a similar extent as the direct effects model, though both support a reliable executive contribution. Given both High and Low Feeling of Knowing measures were based on correct recognition, the additional contribution of executive deficit could be interpreted as the single factor that was associated with the misapplication of the Feeling of Knowing rating. This finding accords with the literature highlighting frontal mediated memory errors, such as confabulation, source memory deficits and high rates of false positive in memory tests (Johnson et al., 2000; Moscovitch & Wincour, 2002; Burgess & Shallice, 1996a). The finding is also more in line with that of Beatty & Monson (1991), and as discussed, may relate to their use of a gamma calculation that reflects absolute, rather than relative, accuracy in that study.

In respect of Feeling of Knowing judgments investigated, it is proposed that executive contributions to metamemory could be summarised as attention to appropriate task features and the evaluation of material that might yield sufficient diagnostic information, notable in prospective judgments, where recall has failed (Schnyer, et al., 2004). The unexpected negative association of information processing with accuracy in Feeling of Knowing judgment might suggest that this probing might have a reflective or intuitive quality, as proposed by Koriat & Goldsmith (1996) and others (Moscovitch & Wincour, 2002), rather than an effortful, stimulus-oriented process. The association with Feeling of Knowing with right hemisphere medial frontal lobe function (Schnyer, et al., 2004) concurs with the hemisphere's proposed role in memory retrieval and subjective evaluation (Tulving et al., 1994), part of which may be the evaluation of familiarity in the presence of degraded recollection (Kelley & Jacoby, 2000).

8.4. A Perspective on Metamemory in Multiple Sclerosis.

Sensitivity to task features

Performance on task-based metamemory measures in the study suggested that the sample demonstrated some sensitivity to task features and demands, evidenced by the underconfidence with practice effect (Koriat, Sheffer & Ma'ayan, 2002) noted across learning trails for the Auditory Verbal Learning Test. Additionally, there was evidence of sensitivity at the first learning trial, with slight overconfidence noted. Both are common findings among unimpaired samples too (Koriat, 2006), suggesting the MS sample is no different from unimpaired samples in this respect.

The second performance component, supporting sensitivity to task difficulty, is the difference in the number of ratings used in the Feeling of Knowing and Retrospective Confidence Judgments in the study. 73% used all 4 ratings in the RCJ task, compared to 17% in the FoK task. Moulin (2002) suggests extent of rating use may be diagnostic of sensitivity to task difficulty in people with Alzheimer's disease. Here in Retrospective Confidence Judgment, range of rating use was consistently greater than on the Feeling of

Knowing task. In the Retrospective Confidence Judgment, the wider use of all four ratings may indicate an ability to more accurately discriminate correctness between retrieved answers. Good discriminative ability in RCJ was supported by the performance indicated by gamma or relative accuracy, which was high for the group as whole, despite varying levels of actual recall.

General Implications

One difficulty with assessing performance of metamemory in a general sense, is that there are limited relationships between the various indices of metamemory themselves. Selective impairments in memory monitoring are a feature of performance of the sample tested here. A difference in accuracy between retrospective and prospective judgments has been documented in a range of studies using clinical samples (Pappas et al., 1992; Schnyer et al., 2004), supporting the lack of a generalised monitoring deficit. Instead it may relate to the facility with which monitoring is achieved on already retrieved, versus unrecallable data. This competence may therefore relate to the use of mnemonic-experience, rather than inferential processes in estimating performance.

This may support differences at the differing levels of awareness presented by Toglia & Kirk (2000) knowledge based, emergent, and anticipatory awareness. Applying these taxonomies would suggest that stored metacognitions about memory, or off-line assessment, may be evidenced through memory self-efficacy measures, implicating perceptions of a participants own memory functioning, beliefs and affective states. Second, emergent awareness, that which develops as a result of experience during tasks, might be characterised in both Judgment of Learning and Retrospective Confidence Judgments. Both generally relate to memory experience diagnostics, including past performance experience in the JoL task and retrieval experience cues in Retrospective Confidence Judgments. Finally, anticipatory awareness captures a range of factors including appraisals of task feature and goal and higher levels of monitoring ability, which might be associated with FoK estimations because of the lack of more concrete experience-related cues from which to form judgments.

The interpretation of Judgment of learning performance, while attributing discrepancies to mnemonic factors, also perhaps implicates task representation failures suggesting some executive contribution. Judgment of Learning therefore, depending on how failure occurs, might equally be considered a mnemonic (a failure to monitor accessibility through covert retrieval) or executive (failure to apprehend the task) phenomenon. If the former, then emergent awareness might be implicated; if the latter, anticipatory awareness might best characterise the area of difficulty. Figure 8.1 summarises one such interpretation, based on differential contribution of mnemonic as opposed to executive processes to judgment accuracy in this study.

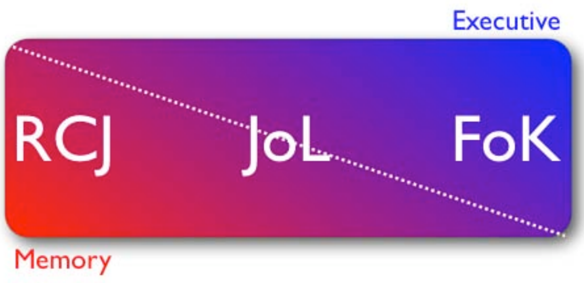


Fig 8.1. Summary of findings suggesting Retrospective Confidence Judgment is primarily mnemonic, and Feeling of Knowing (FoK) primarily executive, with delayed Judgment of learning potentially a combination of mnemonic and executive factors.

Many aspects of current models of metamemory were upheld in the analysis of performance through structural modelling. Retrospective Confidence is primarily mnemonic for episodic tasks, and that Feeling of Knowing judgments are primarily executive in nature. Performance on Judgment of Learning at delay offers less clarity in its interpretation, which may reflect the numerous theories about its dimensions, underconfidence with practice and the delayed Judgment of Learning effect. This study did not evaluate specifically what cues or heuristics are used, so it is not possible to clarify the ongoing debates about information-based or experience-based cues being specifically used in a mnemonic or inferential way. The metamemory literature continues to debate these issues too (Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009). For now however, the presented results can offer some indicators relevant to rehabilitation and learning in people with MS who have, or complain of, memory problems. These include the relevance of mood disorder, the accuracy of post retrieval confidence, the importance of adequate task appraisal and executive function in anticipatory judgments.

A final consideration is a perspective on task complexity; the task themselves offered a continuum from high to low structure and number and type of available cues for making judgments. Variability in both structure and material-richness across the tasks used here suggest that these two factors, if manipulated, offer some opportunities for optimising accuracy. Here, high levels of structure and high levels of memory experience appear to maximise accuracy. In the structured FoK task, assessment of scant cues appear to tap executive weaknesses which may not have been so limiting with increased availability, suggesting some approaches to optimising accuracy. Generally the results suggest that weaknesses in the domain of executive function, because of its relevance in more difficult metamemory monitoring tasks impact more than memory impairment. With memory impairment accuracy can perhaps be achieved, whether calibrated on either the presence or the absence of memory-experience cues. In this sense emergent awareness of memory ability might be considered binary (available, not available) with both the presence and absence of the memory experience being diagnostic.

Anticipatory awareness, relating to a range of judgments not only about the task, but what experiences and cues might best be diagnostic of likely performance, offers less to the judgment maker. In this sense support is available for an extended role of executive abilities is selecting the focus of reflection in respect of the task, rather than just inferring from everything available. This accords with the proposal of Van Overschelde (2008) that while listening in on a telephone call may be an analogy that generally fits with metacognitive monitoring, it is likely to underestimate the active monitoring that takes place. Approaches to improving anticipatory aspects of memory monitoring might therefore focus on remembering as, in and of itself, a purposeful task (Neisser, 1996; Koriat, Goldsmith & Pansky, 2000; Lau, et al., 2004).

8.5. Implications For Memory Rehabilitation And Self-Regulated Learning *in MS.*

In considering the implications for improving memory monitoring in MS, consideration must be given to features of the disease itself, its potential to progress, the lack of a circumscribed locus for damage, meaning that cognitive impairment can be both generalised and incomplete, and the resulting heterogeneity of presentation. The results presented here demonstrate that accuracy in monitoring is likely to be associated with different domains of cognitive function and affect, suggesting intervention should be multidimensional.

Inferring clinical application requires some extrapolation to rehabilitation and daily life; the primary consideration being how the findings might relate to improving self-regulated learning. With this in mind, a conceptual shift is required away from a focus on the accuracy of judgment-making alone to consider the findings of the study within a framework of learning as a task, with specific goals. The monitoring of learning and memory constitutes sub-tasks within that. Relatedly, the meaning of accuracy in memory monitoring should be considered with regard to its ecological validity. From the perspective of social cognition, accuracy has been viewed as correspondence between a judgment and its criterion, as consensus, or in terms of utility (Kruglanski, 1989). While the focus of the study has attended to the correspondence approach, in considering interventions, the additional component of utility becomes important. Baron, (1990) in a commentary on Kruglanski's 1989 paper, suggests that the utility of situational accuracy is its contribution to action and not reflection. It arises from '*veridical contact with the environment*', which becomes the criterion against which the judgment is made (Baron, 1990:201). The task for which an accurate judgment is required may therefore be the true criterion against which accuracy, both in utility and correspondence, is measured.

As the study of metamemory in education, and to a lesser extent in clinical populations, moves towards ecological investigations of monitoring-in-action, so the challenge for clinicians is to consider how monitoring and awareness of memory abilities can be optimised, in relation to the specifics of daily life. It is proposed here that one way of doing this is to focus on accuracy in memory monitoring as part of the skill of memory that contributes much to self-regulated learning. As such, rehabilitation approaches should focus on both metamemory and memory as a process of skill acquisition. These skills might be taught explicitly with a view to them being used procedurally, and generalised, after training.

In addition to this skill focus, Toglia & Kirk (2000:68) suggest that in metacognitive task failures it is often unclear whether inaccurate prediction '*is based on failure to recognise the full demands of the task or based on faulty beliefs about ones abilities*'. In all, three particular dimensions are therefore implicated in considering interventions to optimise the monitoring of memory in self-directed learning; metamemory and related awareness, the task, and the learner. These are now considered in respect of the study's results.

8.5.1. Metamemory and awareness

Attention to memory awareness is important, though is often not the focus when assessment of rehabilitation potential is considered. Instead, object-level measures of cognitive function are often used as indicators of likely rehabilitation success. This seems only partly appropriate in light of the findings of the study, which themselves suggest that there may be dissociations between memory performance and learning abilities, in the sense of being able to accurately appraise even poor memory performance. With accurate appraisal, the successful activation of control operations to manage learning and memory performance is proposed in rehabilitation (Wilson & Moffat, 1992; Prigitano, 1999; Toglia & Kirk, 2000; Clare et al., 2004). The clinical heuristic of patient insight relating to rehabilitation outcome may in part be true, but where insight is assessed by subjective report, it is proposed that only one component of awareness, stored metacognitions about memory, is being considered. This appears from the results of this study to be an index mediated by affective disturbance, unrelated to memory ability.

Given the high levels of depression among people with MS, higher than both the general population and other chronic illness groups (Siegert & Abernethy, 2005), it would be expected that memory complaint will be a common presentation to clinicians during their practice. Findings in respect of memory self-report would therefore suggest that one initial consideration should be of the contribution of mood disorder, with a view to its treatment.

The developing evidence for cognitive-behavioural treatment approaches for depression in MS supports its use for treating mild to moderate depression (Mohr et al., 2000; Mohr & Cox 2001; Mohr, Hart & Goldberg, 2003; Thomas, et al., 2006). Such mild to moderate levels generally accounted for the extent of depression symptom report in this study. The measure of mood disorder used here encompasses more contemporary views of depression as having a basis in negative cognitions (Beck, Steer & Brown, 1996; DSM IV; American Psychiatric Association 2000). This characterisation of depression relates directly to cognitive-behavioural approaches used to treat it, to some effect (Beck, 2002).

Memory-efficacy too may mediate the relationship between mood and learning-oriented behaviours, perhaps explaining more nebulous concepts such as motivation or confidence (Dodds, 1989; Kruglanski, 1989; 1999). Aside from the treatment of mood disorder, a number of studies have investigated whether intervening at the efficacy level would impact on learning effort. Findings suggest that both strategy-use and subsequent performance were positively related to the intervention, including increased maintenance and use of memory strategies (Lachman, et al., 1992; Lachman & Andreoletti, 2006; Lachman, Andreoletti & Pearman, 2006; Dunlosky & Metcalfe, 2009). This is not to suggest that people who had higher domain efficacy improved on measures directly as a result of treatment. Instead, at longer-term follow up, they demonstrated improvements related to their successful use of strategies, which they were more likely to implement (Altmaier et al, 1993). In conjunction with global efficacy-oriented ‘wellness’ programmes, which have demonstrated benefits in MS (Stuifbergen, et al., 2003), there may be scope for developing structured interventions for addressing memory efficacy, in conjunction with management of mood disorder. Such a combined intervention would aim at reducing memory reluctance and increasing deployed effort towards control operations in learning situations.

Studies addressing efficacy therefore suggest it is a malleable construct (Gardiner, Luszcz & Bryan, 1997) and in respect of people with MS, this kind of intervention might be supplemented with education about memory itself. One such educative approach might include promoting understanding that information processing deficits may impact on acquisition, increasing understanding that, although experienced as forgetting, in fact this is likely to relate to failures at acquiring information. Similarly, verbal fluency deficits might be targeted with a similar aim. Once labelled as a memory failure, these too may contribute to overly negative stored metacognitions about memory, which may influence future assessments of memory abilities, selection and deployment of control operations, and effort given to learning.

Aside from approaches that focus on efficacy, the specific monitoring judgments investigated in the study may offer some ways of structuring memory-monitoring tasks to optimise accuracy. Modirrousta & Fellows (2008) suggest that the differences between the Learning, Retrospective Confidence and Feeling of Knowing judgments are their reliance on different weights of various cues and memory strength. This accords with the findings presented; in an MS sample, they are also more or less accurate depending on a mix of cognitive and affective impairment. This suggests that accuracies of the judgments might be dissociable under differing conditions of memory and other cues, with some benefiting from improved memory experiences (Modirrousta & Fellows, 2008). Specifically, training on the skills underlined by the results of this study - internally oriented assessment of the proceeding of memory, evaluation of the task and likely future retrieval demands, the use of covert retrieval, attention to such retrieval experiences as fluency, amount and availability sensations, and training sensitivity to information diagnostic of accuracy - might be approaches taken, concurrent with object level memory rehabilitation.

A final issue relates to the axiomatic use of practice to improve learning and the implications of the development, and unusually, maintenance of underconfidence with practice in the Judgment of Learning result. The question arises as to whether practice may ‘foster underconfidence’ (Koriat, 2002:274) in the calibration of accuracy. Koriat proposes the existence of underconfidence with practice with repetition in action learning, as well as list learning tasks; that is, it is a characteristic of learning generally (Koriat 2002). Here, such calibration in JoL relates to the global assessment of how much is known, potentially in this case leading to continuation of unnecessary of learning, or an efficacy-mediated termination of learning as failing.

8.5.2. The Learner

Discrepancy reduction, between current and desired levels of performance, is one view on the processes involved in learning (Hacker, Bol & Keener, 2008). In memory monitoring, the current level of performance depends in part on knowledge of what is known and not, in order to allocate resources to learning towards reduction of the gap. This supports the importance of these judgments being accurate. The relevance of subject control over resource use to regulate learning is also implied (Son & Kornell, 2008). Extensions to the skill acquisition perspective for metamemory are considerations of how it might be both developed and deployed.

In a sample of undergraduates, the benefits of feedback and training were proposed to be most helpful for better performers, whereas incentive and reward appeared useful for lower achieving students (Hacker, Bol & Keener, 2008). This may be applicable to those with cognitive impairment (Toglia & Kirk, 2000). For those who benefited from feedback, it had to be explicit, and specific to elements of performance (Hacker, Bol & Keener, 2008). Feedback can be graded in terms of from where it comes. Better performers may themselves be able to identify the key feedback elements from within a task. For others, success or failure may be the only usable feedback, meaning that highly structured tasks may be required to give pointers to ongoing performance (Ownsworth et al. 2006; Ownsworth & Clare, 2006). High-level extensions of such approaches to the calibration of judgment accuracy have included the exposure of novices to experts' decision making as a method to improve skilled judgment-making (Karoly, 1993; Weiss, Shanteau & Harries, 2006).

The development of expertise is often also associated with the ability to exclude task-irrelevant information (Karoly, 1993; Dunlosky & Hertzog, 1998). The ability, as is a feature of expertise, requires both experience with a task so as to have access to previous instances of occurrences, but also requires accurate appraisal of a task. With this, comes the deployment of processing resources only towards those features that assist performance or reduce discrepancies (Charness, 1991; Salthouse, 1991; Dunlosky & Bjork, 2008). The task of Judgment of Learning at delay may bear out this latter point. The proposal is that the focus in making the judgment may have been on goal-irrelevant information, previous underconfident judgments, when estimating what was known.

This might support an approach where the modelling of learning-oriented monitoring skills is carried out, including the directing of attention towards information which is diagnostic of accuracy in judgment making. In addition to a focus on diagnosticity of memory or information-based cues, some anticipatory approaches at the learning stage might also improve accuracy in performance prediction. One example of an approach specifically to making the Judgment of Learning is suggested by Nelson, Narens & Dunlosky (2004), the pre-judgment recall and monitoring approach (PRAM); the use of a covert retrieval to make the judgment. A potentially useful elaboration, given the proposed contribution of executive abilities to correctly appraising task contingencies in the first instance (Norman & Shallice, 1986), are interventions focusing on training these kinds of evaluations.

Explicit guidance in appraising ongoing performance, or the nature and goal of the task is proposed to benefit the executive management of ongoing behaviour. Examples of this include training in problem orientation or definition, self-instructional methods to improve monitoring and 'content-free' cueing, to alert to the need for a monitoring episode. All have been shown to be of benefit to people with executive deficits (Cicerone & Wood, 1987; Von Cramon, Mathes-von Cramon & Mai, (1991); Hux, Reid, & Lugert, 1994; Cicerone, et al., 2005; Fisk et al, 2007).

Koriat & Bjork (2005; 2006) also suggest a role for sensitising learners to future retrieval conditions during the learning stage. The results here would suggest a further level of specificity to this recommendation. Sensitivity to the key task demands, and structure, both of which might support anticipatory awareness, and may avoid illusions about competence (Koriat & Bjork, 2005). An approach, which may have utility in this regard, is a focus on correctly defining a task as a memory task, the test portion of which may occur in a supermarket aisle, at a meeting, or during a therapy session.

This issue of appropriate task representation extends, at the learning stage, to attention to the future retrieval context. Echoed from within educational contexts, the advice that students *'study for the type of test you expect to receive'* (Lundberg & Fox, 1991:97 cited in Hacker, Bol & Keener, 2008), emphasises that remembering is a purposive task (Neisser, 1996; Koriat, Goldsmith & Pansky, 2000). In the learning context 'sensitising' goals would relate to envisaging where, and for what purpose, one might be expected to use to-be-learned information, rather than just making a assessment of whether it is remembered or not. This might be considered as emphasising both utility and correspondence views of accuracy. A final point, in respect of generalising monitoring skills between tasks, relates to sensitising learners to the fact that a different learning task may have similarities to other memory tasks in terms of process or monitoring demands (Wykes & Reeder, 2005; Toglia 1991).

8.5.3. The Task

Given the focus on task monitoring contingencies, goals and retrieval context, a task-specific approach to treatment is implicated, because many everyday tasks do not necessarily provide the level of structure or goal clarity that might assist in making accurate monitoring judgments. Adaptation and routinisation of specific tasks may minimise the requirements both for predictive judgment and executive control. General consideration should therefore be given to: grading from high to low levels of structure either by task manipulation, number and type of prompts and monitoring requirements. In this regard, education about task-specific methods of monitoring might be included. Wykes & Reeder, (2005) suggest some additional task-structuring manipulations such as limiting task requirements, the scope of possible responses, managing impulsivity in performance, the use of self-talk during processing as a method of self-regulating performance and emphasising target-oriented, rather than global evaluation. As well as being task-specific, this is also an individualised enterprise because of the centrality of subject-controlled factors in metacognition generally (Koren et al., 2006), and the relationships between awareness, efficacy and knowledge about memory.

A limitation in many studies of metamemory is that participants are not given full control over the learning, meaning that tasks may be more structured than they would be in real-life situations (Koren et al., 2006). This provides for both a positive in terms of task structure maximising performance in those with impairments, and a negative, because without providing latitude for subject-control, the real extent or impact of metamemory deficits may not become apparent. Faglioni et al., (2000) have suggested of people with MS that intentional processes, because they are subject-controlled, may not be affected to the same extent by limitations in information processing speed, perhaps supporting the proposal the key deficit being one of intentional control in the executive relationship with FoK accuracy.

A significant challenge for the future is the development of approaches to evaluating, measuring and structuring such approaches in ecologically relevant ways. Equally, there is a challenge in expanding how clinicians think about memory performance. Shifting from a focus on object-level performance alone to include meta-level processes is required, with the latter including elements of subject control, task dependence and subjective estimation during learning (Koriat, Goldsmith & Pansky, 2000; Koren et al., 2006). The axiom, *‘its not how good your memory is, but what you do with what you have got, that matters’* bears further reflection in this regard, implicating both the monitoring and control processes characterised by the Nelson & Narens’ (1990) model of metacognition.

The metacognitive approach to learning should be an additional, not alternative, approach to structuring assessment and rehabilitation interventions. As yet however, application to clinical intervention has been limited (Koriat, 2002; Moulin, 2002). Future developments in metamemory research in clinical populations will need to involve the development of ecologically valid assessment and the elevation of interventions to everyday learning and memory tasks. A possible direction for the development of measures and interventions is laid out by Koren et al., (2006) in their proposals for a ‘new-approach’ in schizophrenia research. They suggest that metacognition may be able to bridge the gap between object-level cognitive performance and real-life functioning in two key respects; allowing the use of subject-controlled variables to be included as both diagnostic and treatable elements, reflecting real-world instances of individual control over their performance. Second, performance on neuropsychological testing, notably of memory, cannot typically distinguish between the operation of object level memory or metamnemonic processes (Koren et al., 2006). Separation of these components of performance may be of use in assessing whether metacognitive impairments could be used to reduce the ‘rate-limiting’ effects of object level cognitive impairment on learning (Vauth et al., 2005; Koren et al., 2006).

8.6. Appraisal of the study

A number of limitations to the study require acknowledgement, and they relate to three domains specifically; the sample, measurement and analytic strategy. In addition, a number of strengths are indicated, along with avenues for further study.

8.6.1. The sample.

Half of the sample was taking medication for pain, and over a third were prescribed medications for depression, anxiety or sleep disturbance. Many of these medications can have psychoactive effects, and may have a long-term impact on cognitive function (Stewart, 2005). Of interest in this study is whether they have been demonstrated to impact on memory related judgments. A review by Schwartz & Bacon (2008), mainly restricted to benzodiazepine use, is relevant because this class of medications are frequently used to treat muscle spasm, pain, anxiety and insomnia in MS. Summary findings of this review were that measures of relative accuracy were not typically impacted, but more global absolute accuracy was, initially after administration of these drugs. Generally, the findings relate to the acute effects of administration, and so they may be less relevant here, though global assessments in Judgment of Learning and memory efficacy were noted to be inaccurate. Other medications being taken by the group, while representing typical combinations of medications in the population, for pain, spasticity and depression

(Compston et al., 2006), may have had some influence on cognitive performance and affect, both in terms of inhibitory and facilitative impacts (Krupp et al, 2004).

The sample was a convenience sample of community dwelling people with MS. In recruiting the sample, people who had concerns about their memory were sought, and in this sense those who at least had a view on their memory were captured. The aim of the study was to investigate the accuracy of those judgments, and the results of the Memory Function Questionnaire suggest that participants did generally report more problems, with more serious impact, than a standardisation sample. This suggests that the standard of ‘concern about memory’ was met. This is supported by the fact that many did not have memory impairment on objective testing. Participants with more advanced disability that would have been excluded because of motor or sensory impairment may also have been those who were unconcerned about, or unaware of, memory dysfunction and were potentially not captured in this analysis.

One aspect of disease prevalence not fully reflected in the sample was the gender ratio. Just over one fifth of the sample was male. Expected ratios are twice as many women as men. Given that males with MS tend to be diagnosed later, but have poorer outcomes (Tomassini & Pozzilli, 2009), the low recruitment may relate to difficulties accessing more impaired people in a community setting, in addition to a greater propensity to isolation among males with MS (Beatty & Aupperle, 2002; McCabe, McKern & McDonald, 2004). Potentially, these may therefore have been the same people as might be excluded by the criteria relating to fatigue, sensory or motor deficit. Aside from this gender under-representation, comparison with two other UK samples was generally indicative that the sample reported here offers some ability to generalise the findings.

Additional considerations in respect of the sample are individual differences relating to age, gender and intelligence (Heaton, Grant & Matthews, 1986). The study could have been more comprehensive by including an examination of these factors. One example is of the relationships between age and processing speed, and age and executive function (Salthouse, 1996; Verhagen & Salthouse, 1997; Bunce & Macready, 2005 Salthouse, 2010). A possible confound to such an assessment is the extent to which neurological impairment itself may alter the pattern of longitudinal changes in neuropsychological function, so the effects of modeling age would be of interest in the context of other elements. Investigating such age effects should also consider the prevalence of MS subtype and associated patterns of neuropsychological deficit. Generally, older participants might also be those with a longer history of MS and an increased chance of having secondary progressive or primary progressive subtypes, both of which are typically associated with higher levels of cognitive impairment (Beatty et al., 1989; De Sonneville et al., 2002; Fishman et al., 2004), a finding mirrored in this study.

Relatedly, there are considerations appropriate to gender differences in both affect and neuropsychological performance. Evidence from MS samples suggests that males may be more likely to present with more severe cognitive impairment (Tomassini & Pozzilli, 2009). This finding was not investigated here however. An extension is a consideration of gender differences in depression; this could relate to the instrument used or the types of symptoms sought (Salokangas et al., 2002; Ernst & Angst, 1992). One study evaluating depression in a community sample of people with MS suggested that gender was not a factor in prevalence of mood disorder (Chwastiak et al., 2002), but testing associations between gender and depression, and its severity, would still be of interest.

The relationships between intelligence and metamemory also warrant future consideration. A proxy assessment of IQ, such as the National Adult Reading Test (NART; Nelson & Willison, 1991), could probably have been included in the study, without significant costs in terms of fatigue. The main aim of its inclusion would be to extend the analysis of metamemory to include an assessment of the independent contribution of intelligence to accuracy, or to what extent these concepts overlap (Stankov, 2000). Evidence from such studies in education suggests that both intelligence and metacognitive processes exert some unique (independent) effects on performance, disconfirming an intelligence-only model (Veenman & Spaans, 2005; Veenman, Kok & Blote, 2005). Prins, Veenman & Elshout, (2006) suggested that metacognitive skills were most relevant when students were attempting to solve problems at the limits of their current knowledge. The inclusion of an assessment of intelligence in respect of metacognition will also need to define the parameters of the 'intelligence' construct, and the extent to which some aspects, e.g. of fluid intelligence, might be measured by tests of executive function (Duncan, Burgess & Emslie, 1995).

The sample size was large in comparison to most similar studies (Beatty & Monson, 1991; Souchay, Isingrini & Gil, 2002; 2006; Souchay, Bacon & Danion, 2003; Randolph Arnett & Freske, 2004; Perrotin, Belleville & Isingrini, 2007; Julian, Merluzzi & Mohr, 2007), and the level of interest from the target population perhaps attests to its validity as an area of importance. There are additional benefits in a large dataset on a range of measures of object-level cognitive and affective performance, against which future comparison can be made in community-dwelling samples, particularly in the UK. Specifically, the oral only performance of the sample on the Symbol Digit Modalities Test contributes to only a small literature on performance of this test in its less traditional, but potentially more relevant, method of assessment in this population. This is especially relevant because one recent study suggests that it may be a measure with greater utility than the Paced Auditory Serial Addition test, most commonly used to assess processing deficits in MS (Drake et al., 2010). A second range of measures for which data is reported is performance on the Hayling & Brixton tests of executive function. Like the SDMT, the utility of this measure is in its simplicity and brevity as an indicator of executive ability. Confirmatory factor analysis supported it as a measure of this domain, factorially separable from both memory and working memory indices.

Given most 100 participants were seen either in their own homes or at local MS therapy centres, there may be some limitation in respect of a consistent testing environment. This issue of ‘subject-controlled’ variables, implicit in much of metacognitive research could be considered as a confounding factor in equating performance across the sample. Where possible, participants were asked to arrange a quiet environment for the study. Of interest in a number of sessions the researcher did specifically ask that the radio or television be turned off prior to testing, or that a cluttered table might be cleared to avoid distraction. In a clinical context, the requirement for prompting on environmental adaptations might be considered indicative of a failure in participants managing their own learning. This is one area that will be an important consideration for interventional studies. Standardising and allowing subject-controlled variables will offer measurement challenges, but perhaps also offer some insights that might be applied to rehabilitation practice.

From the perspective of the chosen statistical methods, 100 participants would be considered a small, though acceptable number. This means that some aspects of the analysis would benefit from replication in larger samples, specifically the confirmatory factor analysis of the Beck Depression Inventory - II. Statistical fit for these models may have been constrained by the number of parameters being estimated; the degrees of freedom relative to the sample size, being a specific limitation. Related to this interaction of sample size and degrees of freedom are issues of statistical power (Cohen, 1992; Muthen & Muthen, 2002; Kim, 2005; Faul, Erdfelder, Lang & Buchner, 2007).

No formal assessments were carried on the a priori models in determining the sample size required to offer optimum power in the range of fit indices used. A number of approaches are discussed in the literature, with some heuristics informing the discussion of sample size requirements presented in Chapter 3. In reality, the decision about power does not just come down to sample size, but also includes the quality of measures and data, the complexity of models (e.g. degrees of freedom) and the relationships between variables being assessed (Muthen & Muthen, 2002). Two approaches to power assessment in structural equation models are frequently discussed. One uses extrapolation of raw data to derive what would be sufficiently powered parameter estimates, standard errors and confidence intervals, for comparison with the data derived model. This approach includes bootstrapping and Monte Carlo simulations (Methuen & Methuen, 2002; Mooijaart, 2003; Yuan & Hayashi, 2003). Raw data is required, so to some extent is it a post-hoc method of generating power. The other approach relates to testing indices of fit, such as the RMSEA (MacCallum, Browne & Sugawara, 1996; Kline, 2005; Kim, 2005). Both assessments compare hypotheses relating to the power to establish reliable differences between either individual fit statistics (e.g. RMSEA values) or by comparisons of alternative models (MacCallum, Browne & Sugawara, 1996; Kim, 2005). In this study, where many models were accepted because of fit indices and an overall non-significant x^2, the key question relates to the power of the study to be reliable in the values obtained. A formal assessment of power, notably in relation to whether sample size was adequate in the context of degrees of freedom would therefore offer benefit, in addition to the estimation of the size of a future cross-validation sample.

8.6.2. Measurement

Accuracy of the Feeling of Knowing judgment may have been attenuated both by task complexity in the ranking procedure, a feature suggested by a number of authors

(Shimamura & Squire, 1986; Beatty & Monson, 1991; Pannu & Kaszniak, 2005). Being based on an episodic memory task, which may present more difficulty than a semantic task, could be considered a limiting factor on performance (Schwartz & Metcalfe, 1994). One benefit of using a new learning task, aside from its validity in relation to rehabilitation, is in providing equivalence of learning across participants (McKenna & Warrington, 2000). A limitation for the FoK judgment is that magnitude of the gamma relates to the degree of prior learning of the material on which the FoK is based (Nelson et al., 1982). Here, only a single learning trial was given, possibly attenuating the size of the gamma correlation. As mentioned however this may provide a better insight into actual abilities since semantic tasks in MS might not be considered reflective of everyday memory challenges.

In respect of task complexity, comparison of the findings of Souchay, Isingri & Gil, (2002) and Shimamura & Squire, (1986) suggest that the task complexity is less relevant than the level of cognitive deficit. In the former study, only a yes/no response criterion was required among people with Alzheimer's Disease. In the latter a full ranking of items was required, with three cognitive impairment groups. In both studies, as here, mean FoK gammas were found to be near or below zero in the groups of interest, regardless of the task structure. Perhaps therefore, the more relevant finding in respect of task complexity relates to gamma correlations being extracted from an episodic single-exposure task. Though, as proposed by Lad (1984), assessments of subjective probabilities are less related to characteristics of tasks, but rather of the assessor of those probabilities. The difficulty of the task might also mean that it is more sensitive to non-routine, and therefore executive processes. This may indicate that the difficulty of the task rests, in part, with its novelty for participants.

Many of the tasks used might therefore be considered to be novel in the executive sense. They offer limited scope for the application of routine behavioural output to achieve them (Rabbitt, 1997). A simple manipulation for future investigation would be whether accuracy would improve if novelty were reduced through re-testing the same judgments, with the same tasks on the same sample. Of specific interest, would be what level of feedback, if any, might actually be required in improving accuracy.

A second issue in respect of metamemory measurement raised by Spellman, Bloomfield &

Bjork (2008), relates to the amount and quality of information from which the scores were derived (between 2 and 24 unrecalled items in this study), and to the sensitivity of the scale used to classify strength of Feeling of Knowing (initially 1 to 4, followed by rank-sorting within those ratings). Both have been shown to impact on gamma in simulation studies, such as that of Spellman, Bloomfield & Bjork (2008). In a number of studies using similar tasks, additional numbers of items have been used to generate more inclusions of unrecalled items from which to generate potentially more stable indices of accuracy; some include never before seen items also (Beatty & Monson, 1991; Shimamura & Squire, 1986).

A third factor potentially limiting accuracy in the Feeling of Knowing gamma score might relate to the bolstering cue-familiarity experience without concurrently providing exposure to the target sentence completion. Given that a learning trial of all 24 sentences was provided first, and sight of the incomplete sentences given again for the recall trial, and again at the FoK attribution stage, cue-familiarity may have been higher than target familiarity. Resultant from this structure might be a promotion of cue familiarity as a diagnostic cue, but without the commensurate target exposure.

Considerable debate continues in the literature about the appropriateness of Goodman-Kruskal gamma as a measure of FoK accuracy, mainly in respect of its true indication of relative accuracy and its calculation (Goodman & Kruskal 1954; Nelson 1984; Schraw, 1995; Nietfeld, Enders & Schraw 2006; Spellman, Bloomfield & Bjork, 2008). An additional factor from this study relates to the impact of the difficulty of the rating task, indicated by narrow use of even the four-ratings in the FoK task. In terms of calculation of gamma, this difficulty with the FoK rating procedure perhaps led to incomplete rating use and data missingness. Rating non-use meant that cross-tabulated cells, used in the calculation of the Goodman Kruskal gamma correlation, contained zero assignations. The calculation of cross products therefore generated zero scores in either or both numerator and denominators, leading to indeterminate gammas. Souchay, Isingrini & Gil (2006) report a method by which this effect might be negated. It would perhaps be a worthwhile development to investigate if this method provides some equivalence across both rating and ranking methods of calculating gamma.

The issue is specifically addressed in Schraw (1995: citing Freeman, 1987), suggesting that occurrences of empty cells in the matrix is a reason to avoid using the gamma correlation. Summation based-indices, such as the Hamann coefficient have generally not been considered as measures of relative accuracy (Nelson, 1996, Nelson et al., 1986). This rules them out as acceptable alternative indices. It might be expected that on a difficulty task, indicated by the full range of the four ratings not being used, that rating based calculations may therefore present problems, and this might be expected to a greater extent in clinical samples. The full inter-item ranking was used for the calculation of the FoK gamma here and this may be itself a challenge, though it is proposed to capture the relative strength of feeling, avoiding both the limitations of rating non-use and ignoring ties on the rating based procedure.

A second consideration regarding the Goodman Kruskal gamma, notably where some gamma scores are significantly below zero, is the difficulty in interpreting what a near *minus* 1.0 score could mean in relation to a zero score. The latter indicates chance level performance, the former some systematic error in monitoring. Perhaps of particular interest in a clinical context is, assuming task comprehension is ruled out, why someone would be systematically inaccurate. Could these scores relate to specific types of frontal impairment, such as those associated with confabulation, meaning that both unrecognised and recognised are assigned incorrect feelings of knowing? As discussed, many frontal lobe impaired samples do display problems with false confidence (Moscovitch & Wincour, 2002). But, the envisaged performance could only be expected if there was a concomitant memory disorder, which may support the finding that more global cognitive impairment is in fact required to lead to systematic inaccuracy.

A final possibility for systematic inaccuracy relates to the impact of mood disorder on accuracy judgments. This may relate, from the FoK model, to a disposition towards disinhibition and elevated mood, itself associated with significant, often frontal-type cognitive deficit (Cottrell & Wilson, 1926; Finger, 1998; Fishman et al., 2004). A limitation of the study in this regard is not formally assessing both ends of the mood spectrum. Future studies might investigate the euphoric-type mood disturbance in respect of accuracy, and whether there is a reliable association between mood and type of judgments made namely under and over-estimation of accuracy.

A control sample, against which comparisons on performance could be made, might be one method for situating performance of the MS group. This could offer better understanding of the relevance of the range of measurement issues discussed above. While normative data was available for many of the neuropsychological measures used, the metamemory measures, being devised specifically for the purposes of this study, would have benefited from such control data. In addressing the metamemory judgments, submitting control sample performance to similar structural models would also contribute to understanding whether the relationships proposed in the MS sample are more widely applicable.

8.6.3. Analytic Strategy

The proposed benefits of selecting structural equation modelling for this study were the use of latent variables to account for measurement error, to better represent constructs under investigation and to infer some of the mechanisms leading to accuracy and inaccuracy of judgments. However, it is acknowledged that structural equation modelling is a covariance-based method, meaning that associations cannot be interpreted as truly causal mechanisms. This requires many of the findings of this study to be validated in experimental ways. With the extant evidence base, *a priori* models and previous studies these results do offer more than a purely exploratory approach. While *a priori* models were generated, and some rejected, this disconfirmation can only suggest future avenues for research, focusing on potential mechanisms. Additionally, models of mediation presented in the study are speculative at this point because they were in part developed on the basis of earlier results from analysis of the same dataset, and in part from indicators of model modification after initial testing.

A final point on analytic strategy relates to how measures of absolute accuracy were considered, namely by splitting them into accuracy and inaccuracy variables. This approach was motivated by a clinical interest in whether inaccuracy was potentially associated with different processes, compared to accuracy. Perhaps a more careful analysis of the variables would support an interpretation of them reflecting the differences between judgments ascribed as high and low, that is, expectation in the presence of correct recall and recognition. A potential limitation of this approach is that it may lack the *psychological transparency* of measurement, which has been recommended for metamemory measures (Nelson, 1984; Benjamin & Diaz, 2008:77).

Indexing accuracy as proportion of *high ratings and correct* for Retrospective Confidence and Feeling of Knowing, offers clarity into the relationships between predictors and accuracy, but does not offer insight into which participants (e.g. better memory, better executive function) tended to be more accurate in their calibration judgments. These models focus on the contributions to accuracy and inaccuracy. From the models presented there are some indications that the approach was warranted, as comparison between the Feeling of Knowing calibrations suggested that inaccuracy may have been associated with a mediating role for executive function on otherwise sufficient memory ability to achieve recognition.

More standard approaches to absolute accuracy suggest the creation of calibration curves or the generation of an overall calibration score (Hacker, Bol & Keener, 2008). The method use here therefore might be considered to reflect two dimensions of calibration, rather than one; highest and lowest accuracy in the *concordant* High Confidence or FoK, correctly recognised *and* highest and lowest accuracy in *disconcordant* Low Confidence or FoK, correctly recognised. Unlike the other judgments investigated, it is therefore not possible to apply the same between-subject type interpretation with the clarity a single measure might have allowed. Those people who made both high and low judgments are counted in both variables. Their individual balance of High and Low judgments, which would have given an overall assessment of an individual's calibration, are not discernable. Instead, the cognitive contributions to levels of accuracy in each are presented. Of interest in this approach was the failure to find a correlation between the two despite both being calculated as percentage of recalled or recognised targets, which might attest to the two measures being indicative of differing processes, rather than inaccuracy being a failure of accuracy.

Of note, was the ill fitting of the Comparative Fit Index (CFI) in some assessments, less than the ideal > 0.90 value. Rather than being an indicator of the impact of sample size however, CFI values tend to decline as numbers of variables are added to correctly specified models (Kenny & McCoach, 2003) and where there is non-normality in the data (West, Finch & Curran, 1995). It seems appropriate that it was low in the BDI-II CFA, where there were 21 variables assessed and full normality was not achieved, even with transformation of the data.

This is the first reported CFA of the instrument in this population, and confirmed the exploratory analysis of Beck, Steer & Brown (1996). Further validation would benefit future use of the tool in the context of exploring whether one factor is more diagnostic of, or prevalent in depression in MS than the other. An important consideration is whether depression measurement using this tool is confounded with MS symptoms, as proposed in previous versions (Nyenhuis et al., 1995). Resultant from the need to limit the numbers of parameters, factor scores were also created for the two BDI-II dimensions suggested by factor analyses, meaning that measurement error was conflated into their creation. This may explain the occurrence of Heywood cases (negative error variance) occurring in later structural modelling.

8.7. Recommendations for further study

From the results presented, the use of a new learning, single trial task may best capture the problems associated with memory experience as sought in the self-efficacy scale. The association between memory report and acquisition deficits in MS is likely to involve impaired information processing (DeLuca et al, 2004; Lengenfelder et al, 2006; Diamond et al., 2008). Consideration of this finding may demonstrate a more reliable memory-metamemory relationship in future studies. This would be a useful avenue of investigation because there is generally only limited support for memory questionnaires being sensitive to day-to-day memory experience (Hertzog, 2002).

Given the relative novelty of investigations of metamemory in MS, there are a number of elements of this study that would benefit from validation. Perhaps the more significant components are those considered tentative findings here. One of which, relating To information processing and Feeling of Knowing accuracy, presented a finding of particular interest. Further validation is required of the finding that information processing may relate differentially to object-level and meta-level operations. Potentially, this would support a contention than the Feeling of Knowing task, ‘straddles the implicit and explicit’ memory systems (Koriat, 2000).

One factor considered in the discussion of results is the novelty of some of the judgment tasks for participants. In part, this may contribute to the findings themselves, as many of the judgments are considered to have a procedural quality (Reder & Schunn, 1996; Koriat et al., 2008). In this regard, one simple manipulation of interest for a future study would be how performance on any of the metamemory tasks would have changed with repetitions. Having matched unimpaired control data would also support interpretations of performance in the MS group.

One interpretation offered of the lowest scorers on the Beck Depression Inventory was that they might belong to an *absolutely not depressed* group, potentially indicating frontal compromise also (Benedict et al., 2001; Lezak, Howieson & Loring, 2004). For this reason, Mood, rather than depression was considered in this study. However, without a measure of elevated mood, this remains only a tentative conclusion. Indices of disinhibition or euphoric-type mood disorders should be used in future studies so that the impact on accuracy judgments could be evaluated. One general question would be whether depressed mood and cognitive impairment, compared to elevated mood and cognitive impairment might reliably indicate underestimation or overestimation of memory ability. Both of these scenarios might be considered as separate challenges in terms of rehabilitation, self-regulated learning and impact on daily life.

Not fully supported by the findings in this study, but potentially worthy of consideration in advancing understanding of metamemory, heterogeneity in participant performance needs further attention. Beatty & Monson (1991) and Pannu & Kaszniak (2005) make specific recommendations that performance among samples of people with MS might best be considered within functional subgroups so as to best avoid averaging artefacts where there might be clusters of more significant deficit within a sample. This approach would implicate a much larger purposive sample in order to compare the presented models across the different cognitive subgroups, perhaps using a multilevel modelling approach. This does have considerable validity as a recommendation for future studies, especially in cross validating the findings of this study. When the sample was examined based on memory and executive subgroups however, there were no clear differences in the measures of interest that necessarily justified the approach. The structural modelling approach however has the benefit of being able to compare the same modelled interactions for each metamemory index across varying levels of cognitive deficit and affective disturbance. This means it is likely that a cluster of particular variables is likely to associate with an individual metamemory index.

A final recommendation would be for a longitudinal assessment of the models presented in this study, considering whether they remain applicable with cognitive decline or whether the relationships in the model can be adjusted as a result of intervention. Interventional studies should focus on the range of factors highlighted from the study - treatment of mood disorder, recalibration of efficacy, improving memory experience generally, teaching metamemory skills, goal management in the context of everyday memory tasks, task appraisal, and the executive management of task performance. This will necessarily reach into control operations as well as monitoring operations.

8.8. Summary & Conclusions

This study investigated a large UK sample of community-dwelling people with Multiple Sclerosis using a range of metamemory indices, with a focus on monitoring accuracy. The aim of the study was to establish what cognitive and affective factors supported accuracy in the judgments. The measures ranged from off-line memory efficacy to task-based judgments; Judgment of Learning, Retrospective Confidence Judgment and Feeling of Knowing judgment. This range of measures has not been reported before for this population and the scope of the study therefore provides a baseline data for future studies. In this respect, it addresses the call of Pannu & Kaszniak (2005) that a number of metamemory measures be assessed, against a range of measures of cognitive function and mood disorder, in neurological populations.

It is proposed that the range of judgments may generally reflect the three aspects of awareness as proposed by Toglia & Kirk (2000). Knowledge-based awareness, including efficacy and stored metacognitions; emergent awareness, based on engaging in task performance and finally, anticipatory awareness, implicating a representation of the task structure at a more abstract level, implying more executive processes. The movement from post-retrieval judgment, to predictive Judgments of Learning and recognition prediction is proposed to reflect a continuum from memory-supported to executive-supported monitoring judgments.

A global monitoring deficit is not implicated by these findings. Accuracy was associated with memory-supported judgments, regardless of memory ability. This suggests that there are strengths in retrospective confidence and abilities in appraising what is known once a retrieval attempt has been carried out. In a learning prediction task, which may have been achievable through similar post-retrieval assessment, participants apparently failed to use an available store of diagnostic information. Instead, they may have attended to non-diagnostic prior performance indicators, leading to maintained underconfidence in the Judgment of Learning, even after delay. This finding is unusual in the literature, even in clinical groups, and perhaps suggests a failure to attend to factors diagnostic of learning. Finally, in a Feeling of Knowing judgment, executive ability was implicated in accurate performance.

In an extension of the limited metamemory in MS research base, a role for information processing ability was investigated, as it has been supported in facilitating many object-level cognitive processes. Here it was investigated in a similar role in meta-level judgments. Findings offer speculation that contrary to how it facilitates object-level performance in memory, slower information processing speeds were associated with greater Feeling of Knowing accuracy. Once processing speed variance was removed from executive function, a positive relationship between executive function and Feeling of Knowing accuracy emerged. One possibility in understanding this relationship is that the executive processing relating to Feeling of Knowing accuracy is of a reflective, inhibitory nature, rather than one that it deliberative and stimulus oriented. This may concord with an understanding of the Feeling of Knowing judgment as relating more to implicit or reflective memory processes.

A final set of findings relate to the impact of mood disorder on metamemory accuracy. The results suggest that it is the primary contributor to subjective appraisal of memory, including the possibility that it may also bias evaluative contributions of executive function. Additionally it was associated with underconfidence with practice effects in Judgment of Learning over repeated learning trials. Finally, it was reliably related to accuracy in the Feeling of Knowing task. Interpretations for this were discussed as mild depression being associated with better cognitive function in MS. In this sense an artefactual finding, or that mild depression, compared to minimal or none, may, have a facilitating effect on the introspections required in making Feeling of Knowing judgments. A third possibility is that reporting no symptoms of depression in this group could also be an index of elevated mood (Joiner, Schmidt & Metalsky, 2006).

There are a number of avenues for intervention with those who report memory difficulties in MS, relating to improving memory efficacy through the treatment of mood disorder, memory education and offering structured experiences of memory success. Improvements in efficacy are associated with the deployment of memory effort and are therefore of relevance in self-regulated learning. It has been proposed that some potentially more global memory judgments, such as Judgments of Learning might also be considered from an efficacy perspective, potentially explaining the unusual finding of underconfidence with delay in the sample.

What might be classed as executive-related components of performance are most implicated in the study. Specifically, appropriate task apprehension, or selection of relevant diagnostic information may have abolished underconfidence after delay in the Judgment of Learning task. Executive function, and not memory ability, was associated with Feeling of Knowing accuracy. These two results, in conjunction with the high levels of accuracy in retrospective confidence judgments suggest that people with MS who have monitoring deficits should where possible restructure monitoring tasks. Shifting from predictions to postdictions is one obvious recommendation. This might be achieved by using self-testing, covert retrieval or pre-judgment recall and monitoring. These as they are likely to be reliable indicators of what is known, and therefore will support more reliable assessments of performance.

Research in metamemory on clinical populations is at an early stage, perhaps apart from in people with Alzheimer's disease (Moulin, Perfect & Jones, 2000; Moulin, 2002; Moulin, Perfect & Fitch, 2002; Souchay, Isingrini & Gil, 2002; Souchay et al., 2003). There are a number of differences between the two groups that may mean there will be divergence in approaches. People with Alzheimer's disease will continue towards much more significant memory, and cognitive, impairment. Methods to try and negate floor effects in memory performance will need to be specifically developed for some of that group (Moulin, 2002). Secondly, more predictable patterns of decline can be expected in the AD groups meaning findings may be to some extent more applicable. In MS, with notable mood disorder, disinhibition and euphoria, variable cognitive decline, including none, the course of compromised self-directed learning is less predictable. The population is typically of working age and have potential to maintain many functional skills, including employment. An appropriate aim of intervention with this population therefore does relate to self-regulation of learning and performance, making further study of this area warranted both in terms of investigating the nature of deficits and useful interventions.

Novel contribution in this study is first in respect of the sample. This is the first investigation into metamemory performance of people with MS in the United Kingdom. Aside from one postal study (Phillips & Stuifbergen, 2006), which used only self-report measures. This is also the largest MS sample in the literature to investigate metamemory. Additionally, it uses a wide range of measures of metamemory, both self-report and task-based, and assesses them against a number of object-level domains. The findings therefore offer a substantial baseline of performance of this population on many standard indices of metamemory.

The study also provides a number of novel findings in respect of metamemory in Multiple Sclerosis; maintained underconfidence with delay in Judgment of Learning and high levels of accuracy in Retrospective Confidence, despite varying tested memory performance. Results provide confirmation of an executive contribution to Feeling of Knowing accuracy in MS, and suggest this may relate to the top-down directing of processing resources in FoK judgments. Finally, confirmation is offered that in the presence of varying levels of memory ability, including no impairment, people with MS report more memory decline, more memory problems, and consider them more serious when they occur than age matched controls; such memory complaint only being reliably associated with mood.

The study also reports the first confirmatory factor analysis of the revised Beck Depression Inventory (BDI-II) in this population, and the first set of performance data for this population on a 10-item memory self-efficacy scale. The methods used provide a high level of validity to the findings because of the process of variable selection, their factoral confirmation and because measurement error was accounted for in each analysis. Additionally, the testing of theoretically specified priori models confers validity to the found relationships. The presented models extend understanding of metamemory in neurological populations generally, as they provide an approach to thinking about performance in a multivariate way.

Aikens, J., Reinecke, M., Pliskin, N., Fischer, J., Wiebe, J., McCracken, L. & Taylor, J. (1999) 'Assessing depressive symptoms in multiple sclerosis: is it necessary to omit items from the original beck depression inventory?', *Journal of Behavioural Medicine,* 22 (2) pp.127-142.

Akhtar, S., Moulin, C. & Bowie, P. (2006) 'Are people with mild cognitive impairment aware of the benefits of errorless learning?', *Neuropsychological Rehabilitation,* 16 (3) pp.329-346.

Alexander, M. (1994) 'Stroke rehabilitation outcome.' *Stroke,* 25 (1) pp.128-134.

Alexopoulos, G., Meyers, B., Young, R., Kalayam, B., Kakuma, T., Gabrielle, M., Sirey, J. & Hull, J. (2000) 'Executive dysfunction and long-term outcomes of geriatric depression', *Archives Of General Psychiatry,* 57 (3) pp.285-290.

Allison, P. (1987) 'Estimation of linear models with incomplete data', *Sociological Methodology*, 17 pp.71-103.

Altmaier, E., Russell, D. & Lehmann, T. (1993) 'Role of self-efficacy in rehabilitation outcome among chronic low back pain patients', *Journal of Counseling Psychology,* 4 (3) pp.335-339.

American Psychiatric Association (APA). (2000) DSM-IV Diagnostic and Statistical Manual of Mental Disorders. 4th edn. Washington: American Psychiatric Press.

Anderson, D. & Cox, T. (1997) 'Visual signs and symptoms', in: Paty, D. & Ebers, G. (eds) (1998) *Multiple Sclerosis.* New York: Oxford University Press, pp.229-256.

Andrade, J. (ed) (2001a) *Working Memory in Perspective.* Hove: Psychology Press.

Andrade, J. (2001b) 'The contribution of working memory to conscious experience', in: Andrade, J. (ed) *Working Memory in Perspective.* Hove: Psychology Press, pp.60-78.

Andres, P. & Van Der Linden, M. (2000) 'Age-related differences in supervisory attentional system functions', *Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 55B (6) pp.373-380.

Ansell, E. & Bucks, R. (2006) 'Mnemonic anosognosia in Alzheimer's disease: A test of Agnew and Morris (1998)', *Neuropsychologia*, 44 (7) pp.1095-1102.

Arbuckle, J. & Wothke, W. (1999) *Amos 4.0 User's Guide*. Chicago: Smallwaters.

Arbuckle, J. (2007) *Amos 16.0*. Chicago: SPSS Inc.[Software]

Arbuckle, T. & Cuddy, L. (1969) 'Discrimination of item strength at the time of presentation', *Journal of Experimental Psychology,* 81 (1) pp.126-131.

Archibald, C. & Fisk, J. (2000) 'Information processing efficiency in patients with multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 22 (5) pp.686-701.

Archibald, C., Wei, X., Scott, J., Wallace, C., Zhang, Y., Metz, L. & Mitchell, J. (2004) 'Posterior fossa lesion volume and slowed information processing in multiple sclerosis', *Brain*, 127 (7) pp.1526-1534

Arcuri, S., Rabe-Hesketh, S., Morris, R. & Mcguire, P. (2001) 'Regional variation of cloze probabilities for sentence contexts', *Behaviour Research Methods Instruments and Computers,* 33 (1) pp.80-90.

Arnau, R. & Thompson, B. (2000) 'Second order confirmatory factor analysis of the WAIS-III', *Assessment,* 7 (3) pp.237-246.

Arnau, R., Meagher, M., Norris, M. & Bramson, R. (2001) 'Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients', *Health Psychology,* 20 (2) pp.112-119.

Arnett, P. (2004) 'Speed of presentation influences story recall in college students and persons with multiple sclerosis', *Archives of Clinical Neuropsychology,* 19 (4) pp.507-523.

Arnett, P., Higginson, C. & Randolph, J. (2001) 'Depression in multiple sclerosis: relationship to planning ability', *Journal of the International Neuropsychological Society,* 7 (6) pp.665-674.

Arnett, P., Higginson, C., Voss, W., Randolph, J. & Freske, P. (1999a) 'Speed of presentation influences story recall in multiple sclerosis patients', *Archives of Clinical Neuropsychology,* 14 (8) pp.710-711.

Arnett, P., Higginson, C., Voss, W., Randolph, J. & Grandey, A. (2002) 'Relationship between coping, cognitive dysfunction and depression in multiple sclerosis', *The Clinical Neuropsychologist,* 16 (3) pp.341-355.

Arnett, P., Higginson, C., Voss, W., Tippin, J., Clinic, R., Bender, W. & Wurst, J. (1999b) 'Depression in multiple sclerosis: relationship to working memory capacity', *Neuropsychology,* 13 (4) pp.546-556.

Arnett, P., Rao, S., Bernardin, L., Grafman, J., Yetkin, F. & Lobeck, L. (1994) 'Relationship between frontal lobe lesions and Wisconsin Card Sorting Test performance in patients with multiple sclerosis', *Neurology,* 44 (3) pp.420-425.

Arnett, P., Rao, S., Grafman, J., Bernardin, L., Luchetta, T., Binder, J. & Lobeck, L. (1997) 'Executive functions in multiple sclerosis: An analysis of temporal ordering, semantic encoding, and planning abilities', *Neuropsychology,* 11 (4) pp.535-544.

Ascherio, A. & Munger, K. (2007a) 'Environmental risk factors for multiple sclerosis. Part I: The role of infection', *Annals of Neurology,* 61 (4) pp.288-299.

Ascherio, A. & Munger, K. (2007b) 'Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors', *Annals of Neurology,* 61(6) pp.504-513.

Ashworth, B., Aspinall, P. & Mitchell, J. (1989) 'Visual function in multiple sclerosis', *Documenta Ophthalmologica,* 73 (3) pp.209-224.

Asikainen, I., Kaste, M. & Sarna, S. (1998) 'Predicting late outcome for patients with traumatic brain injury referred to a rehabilitation programme: a study of 508 Finnish patients 5 years or more after injury', *Brain Injury,* 12 (2) pp.95-107.

Association of British Neurologists (2009) '*Association of British Neurologists revised guidelines for prescribing in multiple sclerosis',* NHS Evidence - National Library of Guidelines. Available at: http://www.library.nhs.uk/guidelinesFinder/SearchResults.aspx?searchText=association%20of%20british%20neurologists. [Accessed 28 February 2010]

Auger, C., Demers, L. & Swaine, B. (2006) 'Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools', *Archives of Gerontology and Geriatrics,* 43 (1) pp.65-83.

Bacon, E., Danion, J., Kauffmann-Muller, F., Schelstraete, M., Bruant, A., Sellal, F. & Grange, D. (1998) 'Confidence level and feeling of knowing for episodic and semantic memory: an investigation of lorazepam effects on metamemory', *Psychopharmacology,* 138 (3-4) 318-325.

Baddeley, A. (1993) 'A theory of rehabilitation without a model of learning is a vehicle without an engine: a comment on Caramazza and Hillis', *Neuropsychological Rehabilitation,* 3 (3) pp.235-244.

Baddeley, A. (1996) 'The concept of working memory', in: Gathercole, S. (ed) *Models of Short-Term Memory,* Hove: Psychology Press, pp.1-27.

Baddeley, A. (2000) 'The episodic buffer: a new component of working memory?' *Trends in Cognitive Sciences,* 4 (11) pp.417-423.

Baddeley, A., Della Sala, S. & Robbins, T. (1996) 'Working memory and executive control [and discussion]', *Philosophical Transactions: Biological Sciences,* 351 (1346) pp.1397-1404.

Baddeley, A. & Hitch, G. (1974) 'Working Memory', in: Bower, G. (ed) *The Psychology of Learning and Motivation Volume 8.* New York: Academic Press, pp.47-89.

Baddeley, A., & Hitch, G. (2001) 'Foreword', in: Andrade, J. (ed) *Working Memory in Perspective*. 2nd edn. Hove: Psychology Press, pp.xv-xix.

Baddeley, A. & Hitch, G. (2007) ' Working memory: past, present ...and future? in: Osaka, N., Logie, R. & D'esposito, M. (eds) *The Cognitive Neuroscience of Working Memory.* Oxford: Oxford University Press, pp.1-20.

Bahrick, H. (2008) 'Thomas O. Nelson: His life and comments on implications of his functional view of metacognitive memory monitoring', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory*. New York: Psychology Press, pp.1-7.

Bakshi, R. (2003) 'Fatigue associated with multiple sclerosis: diagnosis, impact and management', *Multiple Sclerosis,* 9 (3) pp.219-227

Bakshi, R., Shaikh, Z., Miletich, R., Czarnecki, D., Dmochowski, J., Henschel, K., Janardhan, V., Dubey, N. & Kinkel, P. (2000) 'Fatigue in multiple sclerosis and its relationship to depression and neurologic disability', *Multiple Sclerosis,* 6 (3) pp.181-185

Balota, D., Dolan, P. & Duchek, J. (2000) 'Memory changes in healthy older adults', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory.* New York: Oxford University Press, pp.395-410.

Bandura, A. (1982) 'Self-efficacy mechanism in human agency', *American Psychologist,* 37 (2) pp.122-147.

Bandura, A. (1997) *Self-efficacy: The Exercise of Control*. New York: Worth Publishers.

Barclay, C., Petitto, A., Labrum, A. & Carter-Jessop, L. (1991) 'Mood-related self-schemata and mood-congruity effects in autobiographical memory: A study of women with premenstrual syndrome', *Applied Cognitive Psychology,* 5 (6) pp. 461-481.

Barker-Collo, S. (2005) 'Within session practice effects on the PASAT in clients with multiple sclerosis', *Archives of Clinical Neuropsychology,* 20 (2) pp.145-152.

Baron, R. (1990) 'Kruglanski's lay epistemic theory: when rules describe but do not regulate', *Psychological Inquiry,* 1 (3) pp.200-202.

Baron, R. & Kenny, D. (1986) 'The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations', *Journal of Personality and Social Psychology,* 51 (6) pp.1173-1182.

Barr, W. (2001) ‘Methodologic issues in neuropsychological testing’, *Journal of Athletic Training,* 36 (3) pp.297-302

Barrett, P. (2007) 'Structural equation modelling: Adjudging model fit', *Personality and Individual Differences,* 42 (5) pp.815-824.

Bartlett, F. (1995) *Remembering: A Study in Experimental and Social Psychology*, 2nd edn. reissued. Cambridge: Cambridge University Press.

Bayliss, D., Jarrold, C., Gunn, D. & Baddeley, A. (2003) 'The complexities of complex span: Explaining individual differences in working memory in children and adults', *Journal of Experimental Psychology:General,* 132 (1) pp.71-92.

Beatty, W. (1993) 'Cognitive and emotional disturbances in multiple sclerosis', *Neurologic Clinics,* 11 (1) pp.189-204.

Beatty, W. (1998) 'Cognitive dysfunction in multiple sclerosis: natural history and impact on productive living', *The Italian Journal of Neurological Sciences,* 19 pp.381-387.

Beatty, W. (1999) 'Assessment of cognitive and psychological functions in patients with multiple sclerosis: considerations for databasing', *Multiple Sclerosis,* 5 (4) pp.239-243

Beatty, W. (2004) 'RBANS analysis of verbal memory in multiple sclerosis', *Archives of Clinical Neuropsychology,* 19 (6) pp.825-834.

Beatty, W. & Aupperle, R. (2002) 'Sex differences in cognitive impairment in multiple sclerosis', *The Clinical Neuropsychologist,* 16 (4) pp.472-480.

Beatty, W. & Goodkin, D. (1990) 'Screening for cognitive impairment in multiple sclerosis: an evaluation of the Mini-Mental State Examination', *Archives of Neurology,* 47 (3) pp.297-301.

Beatty, W., Goodkin, D., Beatty, P. & Monson, N. (1989a) 'Frontal lobe dysfunction and memory impairment in patients with chronic progressive multiple sclerosis', *Brain and Cognition,* 11 (1) pp.73-86.

Beatty, W., Goodkin, D., Monson, N. & Beatty, P. (1989b) 'Cognitive disturbances in patients with relapsing remitting multiple sclerosis', *Archives of Neurology,* 46 (10) pp.1113-1119.

Beatty, W., Goodkin, D., Monson, N., Beatty, P. & Hertsgaard, D. (1988) 'Anterograde and retrograde amnesia in patients with chronic progressive multiple sclerosis', *Archives of Neurology,* 45 (6) pp.611-619.

Beatty, W., Hames, K., Blanco, C., Paul, R. & Wilbanks, S. (1995a) 'Verbal abstraction deficit in multiple sclerosis', *Neuropsychology,* 9 (2) pp.198-205.

Beatty, W., Krull, K., Wilbanks, S., Blanco, C., Hames, K. & Paul, R. (1996a) 'Further validation of constructs from the Selective Reminding Test', *Journal of Clinical and Experimental Neuropsychology,* 18 (1) pp.52-55.

Beatty, W. & Monson, N. (1991) 'Metamemory in multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 13 (2) pp.309-327.

Beatty, W., Wilbanks, S., Blanco, C., Hames, K., Tivis, R. & Paul, R. (1996b) 'Memory disturbance in multiple sclerosis: Reconsideration of patterns of performance on the Selective Reminding Test', *Journal of Clinical and Experimental Neuropsychology,* 18 (1) pp.56-62.

Beck, A. (2002) 'Cognitive models of depression', in: Leahy, R. & Thomas Dowd, E. (eds) *Clinical Advances in Cognitive Psychotherapy: Theory and Application.* New York: Springer Publishing, pp29-61.

Beck, A., Steer, R. & Brown, G. (1996) *BDI-II, Beck Depression Inventory: Manual*. 2nd edn. San Antonio: The Psychological Corporation.

Beck, A., Steer, R. & Garbin, M. (1988) 'Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation', *Clinical Psychology Review,* 8 (1) pp.77-100.

Beck, A., Ward, C., Mendelson, M., Mock, J. & Erbaugh, J. (1961) 'An inventory for measuring depression', *Archives of General Psychiatry*, 4 pp.561-571.

Beeney, J. & Arnett, P. (2008) 'Endorsement of self-report neurovegetative items of depression is associated with multiple sclerosis disease symptoms', *Journal of the International Neuropsychological Society,* 14 (6) pp.1057-1062.

Beevers, C. & Miller, I. (2004) ‘Depression-related negative cognition: mood-state and trait dependent properties’, *Cognitive Therapy and Research,* 28**,** pp.293-307.

Benedict, R., Cox, D., Thompson, L., Foley, F., Weinstock-Guttman, B. & Munschauer, F. (2004) ‘Reliable screening for neuropsychological impairment in multiple sclerosis’, *Multiple Sclerosis,* 10 (6) pp.675-678.

Benedict, R., Fischer, J., Archibald, C., Arnett, P., Beatty, W., Bobholz, J., Chelune, G., Fisk, J., Langdon, D. & Caruso, L. (2002) 'Minimal neuropsychological assessment of MS patients: a consensus approach', *The Clinical Neuropsychologist*, 16 (3) pp.381-397.

Benedict, R., Priore, R., Miller, C., Munschauer, F. & Jacobs, L. (2001) 'Personality disorder in multiple sclerosis correlates with cognitive impairment', *Journal of Neuropsychiatry and Clinical Neurosciences*, 13 (1) pp.70-76.

Benjamin, A. & Bjork, R. (1996) 'Retrieval fluency as a metacognitive index', in: Reder, L. (ed) *Implicit Memory and Metacognition.* Mahwah: Lawrence Erlbaum, pp.309-338.

Benjamin, A., Bjork, R. & Schwartz, B. (1998) 'The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index', *Journal of Experimental Psychology-General*, 127 (1) pp.55-68.

Benjamin, A. & Diaz, M. (2008) 'Measurement of relative metamnemonic accuracy', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory*. New York: Psychology Press, pp.73-94.

Bentler, P. & Chou, C. (1987) 'Practical issues in structural modeling', *Sociological Methods & Research*, 16 (1) pp.78-117.

Bentler, P. (1990) 'Comparative fit indexes in structural models', *Psychological Bulletin*, 107 (2) pp.238-246.

Bentler, P. (1994) 'On the quality of test statistics in covariance structure analysis: caveat emptor', in: Reynolds, C. (ed) *Cognitive Assessment: A Multidisciplinary Perspective.* New York: Penum Press, pp.237-260.

Bhatia, K. (1999) 'The paroxysmal dyskinesias', *Journal of Neurology*, 246 (3) pp.149-155.

Bielak, A., Mansueti, L., Strauss, E. & Dixon, R. (2006) 'Performance on the Hayling and Brixton tests in older adults: Norms and correlates', *Archives of Clinical Neuropsychology*, 21 (2) pp.141-149.

Bjartmar, C., Wujek, J. & Trapp, B. (2003) 'Axonal loss in the pathology of MS: consequences for understanding the progressive phase of the disease', *Journal of the Neurological Sciences*, 206 (2) pp.165-171.

Bloom, P. & Fischler, I. (1980) 'Completion norms for 329 sentence contexts', *Memory & Cognition*, 8 (6) pp.631-642.

Blunch, N. (2008) *Introduction to Structural Equation Modelling Using SPSS and AMOS*. London: SAGE Publications.

Boggild, M., Palace, J., Barton, P., Ben-Shlomo, Y., Bregenzer, T., Dobson, C. & Gray, R. (2009) 'Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator', *British Medical Journal*, 339 Accessed online 9.03.10 [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2787922/pdf/bmj.b4677.pdf]

Bogod, N., Mateer, C. & Macdonald, S. (2003) 'Self-awareness after traumatic brain injury: A comparison of measures and their relationship to executive functions', *Journal of the International Neuropsychological Society*, 9 (3) pp.450-458.

Bollen A, & Stine, R. (1993). 'Bootstrapping goodness-of-fit measures in structural equation models', in: Bollen, K. & Long, J. (eds) *Testing Structural Equation Models.* California: Sage Publications, pp.111-135.

Bolker, B. (2008) *Ecological Models And Data In R*. Princeton: Princeton University Press.

Bollen, K. (1989) *Structural Equations With Latent Variables*. New York:Wiley.

Bollen, K. (1990) 'Overall fit in covariance structure models: two types of sample size effects', *Psychological Bulletin*, 107 (2) pp.256-259.

Bollen, K. & Long, J. (eds) (1993) *Testing Structural Equation Models.* California: Sage Publications.

Bollen, K. & Stine, R. (1992) 'Bootstrapping goodness-of-fit measures in structural equation models', *Sociological Methods & Research,* 21 (2) pp.205-229.

Boomsma, A. (1982) 'The robustness of Lisrel against small sample sizes in factor analysis models', in: Joreskog, K. & Wold, H. (eds) *Systems Under Indirect Observation: Causality, Structure, Prediction.* Amsterdam: North Holland

Boomsma, A. & Hoogland, J. (2001) 'The robustness of Lisrel modeling revisited', in: Cudeck, R., Du Toit, S. & Sorbom, D. (eds) *Structural Equation Modeling: Present And Future.* Lincolnwood: Scientific Software Internationl, pp.139-168.

Boringa, J., Lazeron, R., Reuling, I., Ader, H., Pfennings, L., Lindeboom, J., De Sonneville, L., Kalkers, N. & Polman, C. (2001) 'The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice', *Multiple Sclerosis,* 7 (4) pp.263-267.

Boyd, L., Quaney, B., Pohl, P. & Winstein, C. (2007) 'Learning implicitly: effects of task and severity after stroke', *Neurorehabilitation and Neural Repair,* 21 (5) pp.444-454.

Branas, P., Jordan, R., Fry-Smith, A. & Hyde, C. (2000) 'Treatments for fatigue in multiple sclerosis: a rapid and systematic review', *Health Technology Assessment,* 4 (27) pp.1-61.

Brassington, J. & Marsh, N. (1998) 'Neuropsychological aspects of multiple sclerosis', *Neuropsychology Review,* 8 (2) pp.43-77.

Bravin, J., Kinsella, G., Ong, B. & Vowels, L. (2000) 'A study of performance of delayed intentions in multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 22 (3) pp.418-429.

Brewer, W., Sampaio, C. & Barlow, M. (2005) 'Confidence and accuracy in the recall of deceptive and nondeceptive sentences', *Journal Of Memory And Language,* 52 (4) pp.618-627.

Brewer, W. & Sampaio, C. (2006) 'Processes leading to confidence and accuracy in sentence recognition: a metamemory approach', *Memory,* 14 (5) pp.540-552.

Broadbent, D., Cooper, P., Fitzgerald, P. & Parkes, K. (1982) 'The cognitive failures questionnaire (CFQ) and its correlates', *British Journal Of Clinical Psychology,* 21 (1) pp.1-16.

Brooks, N. & Matson, R. (1982) 'Social-psychological adjustment to multiple sclerosis: A longitudinal study', *Social Science and Medicine*, 16 (24) pp.2129-2135.

Brown, R. (1994) 'Efficacy of the indirect approach for estimating structural equation models with missing data: a comparison of five methods', *Structural Equation Modeling: A Multidisciplinary Journal,* 1 (4) pp.287-316.

Brown, T. (2006) *Confirmatory Factor Analysis For Applied Research*. New York: The Guilford Press.

Brown, M. & Cudeck, R. (1993) 'Alternative ways of assessing model fit', in: Bollen, K. & Long, J. (eds) *Testing Structural Equation Models.* Thousand Oaks: Sage Publications, pp.136-162.

Brown, E. & Suppes, T. (1998) 'Mood symptoms during corticosteroid therapy: a review', *Harvard Review of Psychiatry,* 5 (5) pp.239-246.

Brownell, H. & Friedman, O. (2001) 'Discourse ability in patients with unilateral left and right hemisphere brain damage', in: Grafman, J., Boller, F. & Berndt, R. (eds) *Handbook of Neuropsychology, Volume 3 Language and Aphasia.* 2nd Ed Amsterdam: Elservier, pp.189-203.

Brownell, B. & Hughes, J. (1962) 'The distribution of plaques in the cerebrum in multiple sclerosis', *Journal of Neurology, Neurosurgery, and Psychiatry,* 25 (4) pp.315-320.

Bruce, J. & Arnett, P. (2004) 'Self-reported everyday memory and depression in patients with multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 26 (2) pp.200-214.

Bruce, J. & Arnett, P. (2005) 'MS patients with depressive symptoms exhibit affective memory biases when verbal encoding strategies are suppressed', *Journal of the International Neuropsychological Society,* 11 (5) pp.514-521.

Bruce, J., Polen, D. & Arnett, P. (2007) 'Pain and affective memory biases interact to predict depressive symptoms in multiple sclerosis', *Multiple Sclerosis*, 13 (1) pp.58-66.

Buckley, T., Parker, J. & Heggie, J. (2001) 'A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *Journal Of Substance Abuse Treatment,* 20 (3) pp.197-204.

Bugaiska, A., Clarys, D., Jarry, C., Taconnat, L., Tapia, G., Vanneste, S. & Isingrini, M. (2007) 'The effect of aging in recollective experience: The processing speed and executive functioning hypothesis', *Consciousness and Cognition*, 16 (4) pp.797-808.

Bunce, D. (2003) 'Cognitive support at encoding attenuates age differences in recollective experience among adults of lower frontal lobe function', *Neuropsychology,* 17 (3) pp.353-361.

Bunce, D., Handley, R. & Gaines Jr, S. (2008) 'Depression, anxiety, and within-person variability in adults aged 18 to 85 years', *Psychology and Aging,* 23 (4) pp.848-858.

Bunce, D. & Macready, A. (2005) 'Processing speed, executive function, and age differences in remembering and knowing', *The Quarterly Journal of Experimental Psychology Section A,* 58 (1) pp.155-168.

Burgess, P. (1997) 'Theory and methodology in executive function research', in: Rabbitt, P. (ed) *Methodology of Frontal and Executive Function*. Hove: Psychology Press, pp.81-116

Burgess, P. (2000) 'Strategy application disorder: the role of the frontal lobes in human multitasking', *Psychological Research,* 63 (3-4) pp.279-288.

Burgess, P. & Shallice, T. (1996) 'Response suppression, initiation and strategy use following frontal lobe lesions', *Neuropsychologia,* 34 (4) pp.263-272.

Burgess, P. & Shallice, T. (1996a) 'Confabulation and the Control of Recollection', *Memory,* 4 (4) pp.359-411.

Buschke, H. (1973) 'Selective reminding for analysis of memory and learning', *Journal of Verbal Learning and Verbal Behavior,* 12(5) pp.543-550.

Butters, N., Wolfe, J., Granholm, E. & Martone, M. (1986) 'An assessment of verbal recall, recognition and fluency abilities in patients with Huntington's Disease', *Cortex,* 22 (1) pp.11-32.

Byrne, B. (2001) *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming*. Mahwah: Lawrence Erlbaum Associates.

Byrne, B. (2006) *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*. 2nd edn. New York: Routledge.

Calabresi, P. (2004) 'Diagnosis and management of multiple sclerosis', *American Family Physician,* 70 (10) pp.1935-1944.

Calabresi, P. (2006) 'Neuropsychology of multiple sclerosis; an overview', *Journal of Neurology,* 253 (1) pp.i10-i15.

Campbell-Sills, L. & Brown, T. (2005) 'Research considerations: latent variable approaches to studying the classification and psychopathology of mental disorders', in: Andrasik, F., Hersen, M, & Thomas, J. *Comprehensive Handbook of Personality and Psychopathology Volume 2 Adult Psychopathology.* Hoboken: John Wiley.pp.21-35.

Canellopoulou, M. & Richardson, J. (1998) 'The role of executive function in imagery mnemonics: evidence from multiple sclerosis', *Neuropsychologia,* 36 (11) pp.1181-1188.

Cantor, N. & Kihlstrom, J. (1987) *Personality and Social Intelligence*. New Jersey: Prentice Hall.

Carmines, E. & McIver, J. (1981) 'Analyzing models with unobserved variables', in: Bohrnstedt, W. & Borgatta, E. (eds) *Social Measurement: Current Issues,* Beverly Hills: Sage Publications. pp.65-115.

Carone, D., Benedict, R., Munschauer III, F., Fishman, I. & Weinstock-Guttman, B. (2005) 'Interpreting patient/informant discrepancies of reported cognitive symptoms in MS', *Journal of the International Neuropsychological Society,* 11 (5) pp.574-583.

Carroll, M., Gates, R. & Roldan, F. (1984) 'Memory impairment in multiple sclerosis', *Neuropsychologia,* 22 (3) pp.297-302.

Cavanaugh, J. & Green, E. (1990) 'I believe, therefore I can: Self-efficacy beliefs in memory aging', in: Lovelace, E. (ed) *Aging and Cognition: Mental Processes, Self-Awareness, and Interventions.* Amsterdam: Elsevier, pp.189-230.

Cavanaugh, J. & Poon, L. (1989) 'Metamemorial predictors of memory performance in young and older adults', *Psychology and Aging,* 4 (3) pp.365-368.

Charcot, J. (1879) *Lectures on the Diseases of the Nervous System* translated by Sigerson, G. delivered at La Saltpetriere. Charleston: Bibliolife.

Chen, M., Gran, B., Costello, K., Johnson, K., Martin, R. & Dhib-Jalbut, S. (2001) 'Glatiramer acetate induces a Th2-biased response and crossreactivity with myelin basic protein in patients with MS', *Multiple Sclerosis,* 7 (4) pp.209-219.

Chiaravalloti, N., Christodoulou, C., Demaree, H. & Deluca, J. (2003) 'Differentiating simple versus complex processing speed: influence on new learning and memory performance', *Journal of Clinical and Experimental Neuropsychology,* 25 (4) pp.489-501.

Chiaravalloti, N. & Deluca, J. (2008) 'Cognitive impairment in multiple sclerosis', *The Lancet Neurology,* 7 (12) pp.1139-1151.

Chiaravalloti, N., Deluca, J., Moore, N. & Ricker, J. (2005) 'Treating learning impairments improves memory performance in multiple sclerosis: a randomized clinical trial', *Multiple Sclerosis,* 11 (1) pp.58-68.

Chipchase, S. & Lincoln, N. (2001) 'Factors associated with carer strain in carers of people with multiple sclerosis', *Disability & Rehabilitation,* 23 (17) pp.768-776.

Christodoulou, C., Melville, P., Scherl, W., Morgan, T., Macallister, W., Canfora, D., Berry, S. & Krupp, L. (2005) 'Perceived cognitive dysfunction and observed neuropsychological performance: Longitudinal relation in persons with multiple sclerosis', *Journal of the International Neuropsychological Society,* 11 (05) pp.614-619.

Christopher, G. & Macdonald, J. (2005) 'The impact of clinical depression on working memory', *Cognitive Neuropsychiatry,* 10 (5) pp.379-399.

Chua, E., Schacter, D. & Sperling, R. (2009) 'Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments', *Journal of Cognitive Neuroscience,* 21 (9) pp.1751-1765.

Chwastiak, L., Ehde, D., Gibbons, L., Sullivan, M., Bowen, J. & Kraft, G. (2002) 'Depressive symptoms and severity of illness in multiple sclerosis: epidemiologic study of a large community sample', *American Journal of Psychiatry,* 159 (11) pp.1862-1868.

Cicerone, K., Dahlberg, C., Kalmar, K., Langenbahn, D., Malec, J., Bergquist, T., Felicetti, T., Giacino, J., Harley, J. & Harrington, D. (2000) 'Evidence-based cognitive rehabilitation: recommendations for clinical practice', *Archives of Physical Medicine and Rehabilitation,* 81 (12) pp.1596-1615.

Cicerone, K., Dahlberg, C., Malec, J., Langenbahn, D., Felicetti, T., Kneipp, S., Ellmo, W., Kalmar, K., Giacino, J. & Harley, J. (2005) 'Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002', *Archives of Physical Medicine and Rehabilitation,* 86 (8) pp.1681-1692.

Cicerone, K. & Tupper, D. (1991) 'Neuropsychological rehabilitation: treatment of errors in everyday functioning', in: Tupper, D. & Cicerone, K, *The Neuropsychology of Everyday Life: Issues in Development and Rehabilitation.* Norwell: Kluwer Academic Publishers, pp.271-292.

Cicerone, K. & Wood, J. (1987) 'Planning disorder after closed head injury: a case study', *Archives of Physical Medicine and Rehabilitation,* 68 (2) pp.111-115.

Clare, L. (2003) 'Managing threats to self: awareness in early stage Alzheimer's disease', *Social Science & Medicine,* 57 (6) pp.1017-1029.

Clare, L. (2004a) 'Awareness in early-stage Alzheimers disease: A review of methods and evidence', *British Journal of Clinical Psychology,* 43 (2) pp.177-196.

Clare, L. (2004b) 'The construction of awareness in early-stage Alzheimers disease: A review of concepts and models', *British Journal of Clinical Psychology*, 43 (2) pp.155-175.

Clare, L., Wilson, B., Carter, G., Roth, I. & Hodges, J. (2002) 'Assessing awareness in early-stage Alzheimer's disease: Development and piloting of the Memory Awareness Rating Scale', *Neuropsychological Rehabilitation,* 12 (4) pp.341-362.

Clare, L., Wilson, B., Carter, G., Roth, I. & Hodges, J. (2004) 'Awareness in Early-Stage Alzheimers Disease: Relationship to Outcome of Cognitive Rehabilitation', *Journal of Clinical and Experimental Neuropsychology,* 26 (2) pp.215-226.

Clare, L. & Woods, R. (2004) 'Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer s disease: A review', *Neuropsychological Rehabilitation,* 14 (4) pp.385-401.

Clark, D., Beck, A. & Alford, B. (1999) *Scientific Foundations of Cognitive Theory and Therapy of Depression*. New York: John Wiley & Sons.

Clark, C., Fleming, J., Li, D., Oger, J., Klonoff, H. & Paty, D. (1992) 'Sleep disturbance, depression, and lesion site in patients with multiple sclerosis', *Archives of Neurology,* 49 (6) pp.641-643.

Clemmons, D., Fraser, R., Rosenbaum, G., Getter, A. & Johnson, E. (2004) 'An Abbreviated Neuropsychological Battery in Multiple Sclerosis Vocational Rehabilitation: Findings and Implications', *Rehabilitation Psychology,* 49 (2) pp.100-105.

Cohen, R. & Fisher, M. (1989) 'Amantadine treatment of fatigue associated with multiple sclerosis', *Archives of Neurology,* 46 (6) pp.676-680.

Cohen, J (1992) 'A power primer' *Psychological Bulletin,* 112 (1) pp. 115-159.

Cole, J., Motivala, S., Dang, J., Lucko, A., Lang, N., Levin, M., Oxman, M. & Irwin, M. (2004) 'Structural validation of the Hamilton Depression Rating Scale. *Journal of Psychopathology and Behavioral Assessment,* 26 (4) pp.241-254.

Collins, L., Schafer, J. & Kam, C. (2001) 'A comparison of inclusive and restrictive strategies in modern missing data procedures', *Psychological Methods*, 6 (4) pp.330-351.

Compston, A. & Coles, A. (2002) 'Multiple Sclerosis', *The Lancet*, 359 (9313) pp.1221-1231.

Compston, A., Confavreux, H., McDonald, I., Miller, D., Noseworth, J., Kenneth, S. & Wekerle, H. (2006) *McAlpine's Multiple Sclerosis.* 3rd edn. Philadelphia: Churchill Livingstone Elsevier.

Confavreux, A. & Compston. A. (2006) 'The natural history of multiple sclerosis', in: Compson, A., Confavreux, H., McDonald, I., Miller, D., Noseworth, J., Kenneth, S. & Wekerle, H. *McAlpine's Multiple Sclerosis.* 3rd edn. Philadelphia: Churchill Livingstone Elsevier, pp.183-272.

Conklin, L., Strunk, D. & Fazio, R. (2009) 'Attitude formation in depression: Evidence for deficits in forming positive attitudes', *Journal of Behavior Therapy and Experimental Psychiatry,* 40 (1) pp.120-126.

Conway. A., Jarrold, C., Kane., M, Miyake. & Towse, J. (eds) (2007) *Variation in Working Memory*. New York: Oxford University Press.

Cook, S. (ed) (2006) *Handbook of Multiple Sclerosis*. 4th edn. New York: Taylor & Francis Group.

Costelloe, L., Thompson, A., Walsh, C., Tubridy, N. & Hutchinson, M. (2008) 'Long term clinical relevance of criteria for designating Multiple Sclerosis as benign after ten years of disease', *Journal of Neurology, Neurosurgery & Psychiatry*, 79 (11) pp.1245-1248.

Cottrell, S. & Wilson, S. (1926) 'Original papers: The affective symptomatology of disseminated sclerosis: A study of 100 cases', *Journal of Neurology and Psychopathology*, 7 (25) pp.1-30.

Coutler, L. (1989) 'The feeling of knowing in depressed and non-depressed patients with Parkinson's disease', *Journal of Clinical and Experimental Neuropsychology,* 11 pp.91-98.

Cox, D. & Julian, L. (2005) 'Cognitive dysfunction in multiple sclerosis', in: Olek, M. (ed) *Multiple Sclerosis: Etiology, Diagnosis and New Treatment Strategies.* Totowa: Humana Press Inc, pp.91-102.

Coyle, P. (2006) 'Muliple sclerosis biomarkers', in: Cook, S. (ed) (2006) *Handbook of Multiple Sclerosis*. 4th Ed. New York: Taylor & Francis Group, pp.223-242.

Crepeau, E., Willard, H., Cohn, E., Spackman, C. & Schell, B. (2003) 'Occupational Therapy Practice', in: Crepeau, E., Cohn, E. & Boyt Schell, B. (eds) *Willard & Spackman's Occupational Therapy.* 10th edn. Philadephia: Lippincott Williams & Wilkins, pp.27-45

Daly, L. & Bourke, G. (eds) (2000) *Interpretation and Uses of Medical Statistics*. 5th edn. Oxford: Blackwell Science.

Daneman, M. & Hannon, H. (2007) 'What do working memory span tasks like reading span really measure?', in: Osaka, N., Logie, R. & D'esposito, M. (eds) *The Cognitive Neuroscience of Working Memory.* Oxford: Oxford University Press, pp.21-42.

Davis, J. (1967) 'A partial coefficient for Goodman and Kruskal's Gamma', *Journal of the American Statistical Association*, 62 (317) pp.189-193.

De Frias, C., Dixon, R. & Strauss, E. (2006) 'Structure of four executive functioning tests in healthy older adults', *Neuropsychology,* 20 (2) pp.206-214.

De Sonneville, L., Boringa, J., Reuling, I., Lazeron, R., Ader, H. & Polman, C. (2002) 'Information processing characteristics in subtypes of multiple sclerosis', *Neuropsychologia,* 40 (11) pp.1751-1765.

Delaney, P., Reder, L., Staszewski, J. & Ritter, F. (1998) 'The strategy-specific nature of improvement: The power law applies by strategy within task', *Psychological Science,* 9 (1) pp.1-7.

Delis, D. & Kramer, J. (2000) 'Advances in the Neuropsychological Assessment of Memory Disorders', in: Cermak, L. (ed) *Handbook of Neuropsychology Volume 2 Memory and its Disorders.* 2nd edn. Amsterdam: Elsevier, pp.25-48.

Delis, D., Kramer, J., Kaplan, E. & Ober, B. (1987) *California Verbal Learning Test: Adult Version*. San Antonio: The Psychological Corporation.

Deloire, M., Bonnet, M., Salort, E., Arimone, Y., Boudineau, M., Petry, K. & Brochet, B. (2006) 'How to detect cognitive dysfunction at early stages of multiple sclerosis?, *Multiple Sclerosis,* 12 (4) pp.445-452.

DeLuca, J. (2006) 'What we know about cognitive changes in multiple sclerosis', in: LaRocca, N. & Kalb, R. (eds) *Multiple Sclerosis: Understanding the Cognitive Challenges*. New York: Demos Medical Publishing, pp.17-40.

DeLuca, J., Chelune, G., Tulsky, D., Lengenfelder, J. & Chiaravalloti, N. (2004a) 'Is processing speed or working memory the primary information processing deficit in multiple sclerosis?' *Journal of Clinical and Experimental Neuropsychology*, 26 (4) pp.550-562.

DeLuca, J., Christodoulou, C., Diamond, B., Rosenstein, E., Kramer, N. & Natelson, B. (2004b) 'Working memory deficits in chronic fatigue syndrome: Differentiating between speed and accuracy of information processing', *Journal of the International Neuropsychological Society,* 10 (1) pp.101-109.

Demaree, H., Deluca, J., Gaudino, E. & Diamond, B. (1999) 'Speed of information processing as a key deficit in multiple sclerosis: implications for rehabilitation', *Journal of Neurology, Neurosurgery & Psychiatry*, 67 (5) pp.661-663.

Demaree, H., Gaudino, E. & Deluca, J. (2003) 'The relationship between depressive symptoms and cognitive dysfunction in multiple sclerosis', *Cognitive Neuropsychiatry,* 8 (3) pp.161-171.

Denney, D., Lynch, S., Parmenter, B. & Horne, N. (2004) 'Cognitive impairment in relapsing and primary progressive multiple sclerosis: mostly a matter of speed', *Journal of the International Neuropsychological Society,* 10 (7) pp.948-956.

Denney, D., Sworowski, L. & Lynch, S. (2005) 'Cognitive impairment in three subtypes of multiple sclerosis', *Archives of Clinical Neuropsychology,* 20 (8) pp.967-981.

Department Of Health, UK (2002). Health Service Circular 2002/004. '*Cost effective provision of disease modifying therapies for people with multiple sclerosis*', Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Healthservicecirculars/DH_4004332 [accessed 28/2/2010]

Department Of Health, UK (2002). '*Drug treatments for multiple sclerosis (beta-interferons and glatiramer acetate): risk sharing scheme (HSC 2002/004) briefing note November 2006'*, Available at*:* http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_063445 [accessed 28/2/2010]

DeSouza, L. & Bates, D. (2004) 'Multiple Sclerosis' in: Stokes, M. (Ed.) *Physical Management in Neurological Rehabilitation*. 2nd edn. London: Elsevier Mosby, pp.177-202.

Devolder, P. & Pressley, M. (1992) 'Causal attributions and strategy use in relation to memory performance differences in younger and older adults', *Applied Cognitive Psychology,* 6 (7) pp.629-642.

Di Fabio, R. Soderberg J, Choi T, Hansen CR, Shapiro RT. (1998) 'Extended outpatient rehabilitation: its influence on symptom frequency, fatigue, and functional status for persons with progressive multiple sclerosis', *Archives of Physical Medicine and Rehabilitation*, 79 (2) pp.141-146.

Diamond, B., Deluca, J., Johnson, S. & Kelley, S. (1997a) 'Verbal learning in amnesic anterior communicating artery aneurysm patients and in patients with multiple sclerosis', *Applied Neuropsychology*, 4 (2) pp.89-98.

Diamond, B., Deluca, J., Kim, H. & Kelley, S. (1997b) 'The question of disproportionate impairments in visual and auditory information processing in multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology*, 19 (1) pp.34-42.

Diamond, B., Johnson, S., Kaufman, M. & Graves, L. (2008) 'Relationships between information processing, depression, fatigue and cognition in multiple sclerosis', *Archives of Clinical Neuropsychology*, 23 (2) pp.189-199.

Diamond, P., Felsenthal, G., Macciocchi, S., Butler, D. & Lally-Cassady, D. (1996) 'Effect of Cognitive Impairment on Rehabilitation Outcome', *American Journal of Physical Medicine & Rehabilitation*, 75 (1) pp.40-43.

Ding, L., Velicer, W. & Harlow, L. (1995) 'Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices', *Structural Equation Modeling: A Multidisciplinary Journal*, 2 (2) pp.119-143.

Dixon, R. & Hultsch, D. (1983) 'Structure and development of metamemory in adulthood', *The Journal of Gerontology*, 38 (6) pp.682-688.

Dixon, R. & Hultsch, D. (1984) 'The metamemory in adulthood (MIA) instrument', *Psychological Documents*, 14 (3).

Dixon, R., Hultsch, D. & Hertzog, C. (1988) 'The Metamemory in Adulthood (MIA) questionnaire', *Psychopharmacology Bulletin*, 24 (4) pp.671-688.

Dodds, A. (1989) 'Motivation reconsidered: the importance of self-efficacy in rehabilitation', *British Journal of Visual Impairment*, 7 (1) pp.11-15.

Dozois, D., Dobson, K. & Ahnberg, J. (1998) 'A psychometric evaluation of the Beck Depression Inventory-II', *Psychological Assessment*, 10 (2) pp.83-89.

Drake, A., Weinstock-Guttman, B., Morrow, S., Hojnacki, D., Munschauer, F. & Benedict, R. (2010) 'Psychometrics and normative data for the Multiple Sclerosis Functional Composite: replacing the PASAT with the Symbol Digit Modalities Test', *Multiple Sclerosis*, 16 (2) pp.228-237.

Drew, M., Starkey, N. & Isler, R. (2009) 'Examining the Link between Information Processing Speed and Executive Functioning in Multiple Sclerosis', *Archives of Clinical Neuropsychology*, 24 (1) pp.47-58

Duncan, J., Burgess, P. & Emslie, H. (1995) 'Fluid intelligence after frontal lobe lesions', *Neuropsychologia*, 33 (3) pp.261-268.

Duncan, J., Johnson, R., Swales, M. & Freer, C. (1997) 'Frontal lobe deficits after head injury: Unity and diversity of function', *Cognitive Neuropsychology,* 14 (5) pp.713-741.

Dunlosky, J. & Bjork, R. (eds) (2008a) *Handbook of Memory and Metamemory.* New York: Psychology Press.

Dunlosky, J. & Bjork, R. (2008b) 'The integrated nature of metamemory and memory', in: Dunlosky, J. & Bjork, R. *Handbook of Metamemory and Memory.* New York:Psychology Press, pp.11-28.

Dunlosky, J., Hertzog, C., Kennedy, M. & Thiede, K. (2005) 'The Self-Monitoring Approach For Effective Learning', *International Journal of Cognitive Technology,* 10 (1) pp.4-11.

Dunlosky, J., Kubat-Silman, A. & Hertzog, C. (2003) 'Training monitoring skills improves older adults' self-paced associative learning', *Psychology and Aging*, 18 (2) pp.340-345.

Dunlosky, J. & Metcalfe, J. (2009) *Metacognition*. Los Angelas: Sage Publications.

Dunlosky, J. & Nelson, T. (1992) 'Importance of the kind of cue for judgments of learning (Jol) and the delayed-Jol effect', *Memory & Cognition,* 20 (4) pp.374-380.

Dunlosky, J., Rawson, K. & Mcdonald, S. (2002) 'Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material', in: Perfect, T. & Schwartz, B. *Applied Metacognition.* Cambridge: Cambridge University Press, pp.68-94.

Dunlosky, J., Serra, M. & Baker, J. (2007) 'Metamemory', in: Durso, F., Nickerson, R., Dumais, S., Lewandowsky, S. & Perfect, T. (eds). 2nd edn. *Handbook of Applied Cognition.* Hoboken: Wiley, pp.137-160.

Durso, F., Nickerson, R., Dumais, S., Lewandowsky, S & Perfect, T. (eds) (2007) *Handbook of Applied Cognition 2nd edn* Chichester: John Wiley.

Ebers, G. (2001) 'Natural History Of Multiple Sclerosis', *Journal of Neurology, Neurosurgery, and Psychiatry,* 71 (Suppl 2) ppe.16-19.

Ebers, G. (2008) 'Environmental factors and multiple sclerosis', *The Lancet Neurology,* 7 (3) pp.268-277.

Efklides, A. (2006) 'Metacognitive Experiences: The Missing Link in the Self-Regulated Learning Process', *Educational Psychology Review*, 18 (3) pp.287-291.

Efron, B. (1979) 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics,* 7 (1) pp.1-26.

El-Moslimany, H. & Lublin, F. (2008) 'Clinical features in multiple sclerosis', in: Raine, C., Mcfarland, H. & Hohlfeld, R. *Multiple Sclerosis: A Comprehensive Text*. London: Saunders Elsevier, pp.10-23.

Elliott, E. & Lachman, M. (1989) 'Enhancing memory by modifying control beliefs, attributions, and performance goals in the elderly', in: Fry, P. (ed) *Psychological Perspectives of Helplessness and Control in the Elderly*. Amsterdam: Elsevier, pp.339-367.

Eraksoy, M. (1999) 'Multiple sclerosis in children: a review', in: Siva, A., Kesselring, J. & Thompson, J. (eds) *Frontiers in Multiple Sclerosis Volume 2*. London: Martin Dunitz, pp.67-73.

Ernst, C & Angst, J (1992) 'The Zurich Study: XII. Sex differences in depression. Evidence from longitudinal epidemiological data. *Psychiatry and Clinical Neuroscience*, 241 pp. 222-230.

Faglioni, P., Bertolani, L., Botti, C. & Merelli, E. (2000) 'Verbal learning strategies in patients with multiple sclerosis', *Cortex*, 36 (2) pp.243-263.

Faul, F., Erdfelder, E., Lang, A-G. & Buchner, A. (2007) 'G*Power: a flexible statistical power analysis for the social, behavioural and biomedical sciences' *Behavioural research Methods*, 39 (2) pp. 175-191.

Feinstein, A. (2007) *The Clinical Neuropsychiatry of Multiple Sclerosis*, 2nd edn. Cambridge: Cambridge University Press.

Feinstein, A., O'Connor, P. & Feinstein, K. (2002) 'Multiple Sclerosis, interferon beta-1b and depression', *Journal of Neurology,* 249 (7) pp.815-820.

Feinstein, A., Roy, P., Lobaugh, N., Feinstein, K., O'Connor, P. & Black, S. (2004) 'Structural brain abnormalities in multiple sclerosis patients with major depression', *Neurology,* 62 (4) pp.586-590.

Fidell, L. & Tabachnick, B. (2003) 'Preparatory data analysis', in: Schinka, J., Velicer, W. (eds) *Handbook of Psychology: Volume 2 Research Methods in Psychology*. Hoboken: John Wieley & Sons pp.114-141.

Field, A. (2005) Discovering Statistics using SPSS. 2nd edn. London: Sage Publications.

Finger, S. (1998) 'A happy state of mind: a history of mild elation, denial of disability, optimism, and laughing in multiple sclerosis', *Archives of Neurology,* 55 (2) pp.241-250.

Finlayson, M., Garcia, J. & Cho, C. (2008) 'Occupational therapy service use among people aging with multiple sclerosis', *American Journal of Occupational Therapy,* 62 (3) pp.320-328.

Finn, B. & Metcalfe, J. (2007) 'The role of memory for past test in the underconfidence with practice effect', *Journal Of Experimental Psychology Learning Memory and Cognition,* 33 (1) pp.238-244.

Finn, B. & Metcalfe, J. (2008) 'Judgments of learning are influenced by memory for past test', *Journal of Memory and Language,* 58 (1) pp.19-34.

Finney, S. & DiStefano, C. (2006) 'Non-normal and categorical data in structural equation modeling', in: Hancock, G. & Mueller, R. (eds) *Structural Equation Modeling: A Second Course.* Greenwich: Information Age Publishing, pp.269-314.

Fischer, J. (2001) 'Cognitive impairment in multiple sclerosis', in: Cook, S. (ed) *Handbook of Multiple Sclerosis*. 3rd edn. New York: Taylor & Francis group, pp.233-255.

Fischer, S., Gauggel, S. & Trexler, L. (2004) 'Awareness of activity limitations, goal setting and rehabilitation outcome in patients with brain injuries', *Brain Injury,* 18 (6) pp.547-562.

Fishman, I., Benedict, R., Bakshi, R., Priore, R. & Weinstock-Guttman, B. (2004) 'Construct validity and frequency of euphoria sclerotica in multiple sclerosis', *Journal of Neuropsychiatry and Clinical Neurosciences,* 16 (3) pp.350-356.

Fisk, J. & Archibald, C. (2001) 'Limitations of the Paced Auditory Serial Addition Test as a measure of working memory in patients with multiple sclerosis', *Journal of the International Neuropsychological Society,* 7 (3) pp.363-372.

Fiszdon, J., Mcclough, J., Silverstein, S., Bell, M., Jaramillo, J. & Smith, T. (2006) 'Learning potential as a predictor of readiness for psychosocial rehabilitation in schizophrenia', *Psychiatry Research,* 143 (2) pp.159-166.

Flavell, J. (1971) 'Stage-related properties of cognitive development', *Cognitive Psychology,* 2 (4) pp.421-453.

Flavell, J. (1979) 'Metacognition and cognitive monitoring: A new area of cognitive-development', *American Psychologist,* 34 (10) pp.906-911.

Flavell, J., Friedrichs, A. & Hoyt, J. (1970) 'Developmental changes in memorization processes', *Cognitive Psychology,* 1 (4) pp.324-340.

Fletcher, P., Shallice, T. & Dolan, R. (2000) ''Sculpting the Response Space'' - An Account of Left Prefrontal Activation at Encoding', *Neuroimage,* 12 (4) pp.404-417.

Fidell, L. & Tabachnick, B. (2003) 'Preparatory data analysis' in: Schinka, J. & Velicer, W. (eds) *Handbook of Psychology: Volume 2 Research Methods in Psychology*. Hoboken: John Wiley & Sons pp.114-141.

Floyd, F. & Widaman, K. (1995) 'Factor analysis in the development and refinement of clinical assessment instruments', *Psychological Assessment,* 7 (3) pp.286-299.

Foley, G. (2008) 'Occupational Therapy in Progressive Neurology: a Rehabilitative Approach', *The British Journal of Occupational Therapy,* 71 (7) pp.308-310.

Ford, H., Gerry, E., Airey, C., Vail, A., Johnson, M. & Williams, D. (1998) 'The prevalence of multiple sclerosis in the Leeds Health Authority', *Journal of Neurology, Neurosurgery & Psychiatry,* 64 (5) pp.605-610.

Ford, A. & Naismith, R. (2006) 'Depression and multiple sclerosis', in: Gilliam, F., Kanner, A. & Sheline, Y. (eds) *Depression and Brain Dysfunction.* New York: Taylor & Francis, pp.263-270.

Foreman, M., Fletcher, K., Mion, L. & Trygstad, L. (2003) 'Assessing cognitive function', in: Mezey, M., Fulmer, T. & Abraham, I. *Geriatric Nursing Protocols for Best Practice 2nd edn.* New York: Springer, pp.99-114.

Fos, L., Greve, K., South, M., Mathias, C. & Benefield, H. (2000) 'Paced Visual Serial Addition Test: an alternative measure of information processing speed', *Applied Neuropsychology,* 7 (3) pp.140-146.

Fouladi, R. (1998) 'Covariance structure analysis techniques under conditions of multivariate normality and nonnormality-modified and bootstrap based test statistics', *Paper presented at the annual meeeting of the American Educational Research Association, San Diego, Califonia.*

Freal, J., Kraft, G. & Coryell, J. (1984) 'Symptomatic fatigue in multiple sclerosis', *Archives of Physical Medicine and Rehabilitation,* 65 (3) pp.135-138.

Frith, C. & Frith, U. (1999) 'Interacting minds-a biological basis', *Science,* 286 (5445) pp.1692-1695.

Fritz, M. & Mackinnon, D. (2007) 'Required sample size to detect the mediated effect', *Psychological Science,* 18 (3) pp.233-239.

Fu, T., Koutstaal, W., Fu, C., Poon, L. & Cleare, A. (2005) 'Depression, confidence, and decision: Evidence against depressive realism', *Journal of Psychopathology and Behavioral Assessment,* 27 (4) pp.243-252.

Fuster, J. (2002) 'Physiology of executive functions: The perception-action cycle', in: Stuss, D. & Knight, R. (eds) *Principles of Frontal Lobe Function,* Oxford: Oxford University Press, pp.96-108.

Gale, C. & Martyn, C. (1995) 'Migrant studies in multiple sclerosis', *Progress in Neurobiology,* 47 (4-5) pp.425-448.

Galski, T., Bruno, R., Zorowitz, R. & Walker, J. (1993) 'Predicting length of stay, functional outcome, and aftercare in the rehabilitation of stroke patients. The dominant role of higher-order cognition', *Stroke,* 24 (12) pp.1794-1800.

Garavan, H., Ross, T., Murphy, K., Roche, R. & Stein, E. (2002) 'Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction', *Neuroimage,* 17 (4) pp.1820-1829.

Gardiner, M., Luszcz, M. & Bryan, J. (1997) 'The manipulation and measurement of task-specific memory self-efficacy in younger and older adults', *International Journal of Behavioral Development,* 21 (2) pp.209-227.

Giacino, J. & Cicerone, K. (1998) 'Varieties of deficit unawareness after brain injury', *Journal of Head Trauma Rehabilitation,* 13 (5) pp.1-15.

Gibbons, J. (1993) *Nonparametric Measures of Association*. Newbury Park: Sage Publications.

Gilewski, M. & Zelinski, E. (1986) 'Questionnaire assessment of memory complaints', in: Poon, L. (ed) *Handbook for Clinical Memory Assessment of Older Adults,* Washington DC: American Psychological Association, pp.93-107.

Gilewski, M. & Zelinski, E. (1988) 'Memory Functioning Questionnaire (MFQ)', *Psychopharmacology Bulletin,* 24 (4) pp.665-670.

Gilewski, M., Zelinski, E. & Schaie, K. (1990) 'The Memory Functioning Questionnaire for assessment of memory complaints in adulthood and old age', *Psychology and Aging,* 5 (4) pp.482-490.

Giovannoni, G. & Ebers, G. (2007) 'Multiple sclerosis: the environment and causation', *Current Opinion in Neurology,* 20 (3) pp.261-268.

Gironi, M., Martinelli-Boneschi, F., Sacerdote, P., Solaro, C., Zaffaroni, M., Cavarretta, R., Moiola, L., Bucello, S., Radaelli, M. & Pilato, V. (2008) 'A pilot trial of low-dose naltrexone in primary progressive multiple sclerosis', *Multiple Sclerosis,* 14 (8) pp.1076-1083.

Gisiger, T., Kerszberg, M. & Changeux, J. (2005) 'Acquisition and performance of delayed-response tasks: a neural network model', *Cerebral Cortex,* 15 (5) pp.489-506.

Glenberg, A. & Epstein, W. (1987) 'Inexpert calibration of comprehension', *Memory & Cognition,* 15 (1) pp.84-93.

Glisky, E. & Glisky, M. (2002) 'Learning and memory impairments', in: Eslinger, P. (ed) '*Neuropsychological Interventions: Clinical Research and Practice.* New York: Guilford Press, pp.137-162.

Godefroy, O. (2003) 'Frontal syndrome and disorders of executive functions', *Journal of Neurology,* 250 (1) pp.1-6.

Goel, V. & Grafman, J. (1995) 'Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi', *Neuropsychologia,* 33 (5) pp.623-642.

Goel, V., Pullara, D. & Grafman, J. (2001) 'A computational model of frontal lobe dysfunction: working memory and the Tower of Hanoi task', *Cognitive Science,* 25 (2) pp.287-313.

Gold, S. & Irwin, M. (2006) 'Depression and immunity: inflammation and depressive symptoms in multiple sclerosis', *Neurologic Clinics,* 24 (3) pp.507-519.

Goldberg, E. (2001) *The Executive Brain: Frontal Lobes and the Civilized Mind*. Oxford: Oxford University Press.

Goldman Consensus Group (2005) 'The Goldman Consensus group statement on depression in multiple sclerosis', *Multiple Sclerosis,* 11 (3) pp.328-337.

Goldman-Rakic, P. (1994) 'Working memory dysfunction in schizophrenia', *Journal of Neuropsychiatry & Clinical Neurosciences,* 6 (4) pp.348-357.

Goldsmith, M., Koriat, A. & Pansky, A. (2005) 'Strategic regulation of grain size in memory reporting over time', *Journal of Memory and Language,* 52 (4) pp.505-525.

Gontkovsky, S. & Beatty, W. (2006) 'Practical methods for the clinical assessment of information processing speed', *International Journal of Neuroscience,* 116 (11) pp.1317-1325.

Goodman, L. & Kruskal, W. (1954) 'Measures of association for cross classifications', *Journal of the American Statistical Association* 49 pp.732–764.

Goodman, L. & Kruskal, W. (1959) 'Measures of association for cross classifications. II: further discussion and references', *Journal of the American Statistical Association,* 54 (285) pp.123-163.

Gotlib, I., Roberts, J. & Gilboa, E. (1996) 'Cognitive interference in depression', in: Sarason, I., Pierce, G. & Sarason, B. *Cognitive Interference: Theories, Methods, and Findings*. Mahwah. Lawrence Erlbaum Associates.

Goverover, Y., Chiaravalloti, N. & Deluca, J. (2005) 'The relationship between self-awareness of neurobehavioral symptoms, cognitive functioning, and emotional symptoms in multiple sclerosis', *Multiple Sclerosis,* 11 (2) pp.203-212.

Goverover, Y., Chiaravalloti, N. & Deluca, J. (2008a) 'Self-generation to improve learning and memory of functional activities in persons with multiple sclerosis: meal preparation and managing finances', *Archives of Physical Medicine and Rehabilitation,* 89 (8) pp.1514-1521.

Goverover, Y., Genova, H., Hillary, F. & Deluca, J. (2007) 'The relationship between neuropsychological measures and the Timed Instrumental Activities of Daily Living task in multiple sclerosis', *Multiple Sclerosis,* 13 (5) pp.636-644.

Goverover, Y., Hillary, F., Chiaravalloti, N., Arango-Lasprilla, J. & Deluca, J. (2008b) 'A functional application of the spacing effect to improve learning and memory in persons with multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 31 (5) pp.513-522.

Grafman, J., Rao, S., Bernardin, L. & Leo, G. (1991) 'Automatic memory processes in patients with multiple sclerosis', *Archives of Neurology,* 48 (10) pp.1072-1075.

Grafman, J., Rao, S. & Litvan, I. (1990) 'Disorders of memory', in: Rao, S, (ed) *Neurobehavioral Aspects of Multiple Sclerosis,* Oxford: Oxford University Press, pp102-117.

Graham, J. (2009) 'Missing data analysis: making it work in the real world', *Annual Review Of Psychology,* 60 pp.549-576.

Graham, J., Cumsille, P. & Elek-Fisk, E. (2003) 'Methods for handling missing data', in: Schinka, J. & Velicer, W. (eds) *Handbook Of Psychology: Research Methods in Psycholoy.* Hoboken: John Wiley, pp.87-114.

Grant, D. & Berg, E. (1948) *The Wisconsin Card Sort Test: Directions for Administration and Scoring. Odessa: Psychological Assessment.*

Grasso, M., Troisi, E., Rizzi, F., Morelli, D. & Paolucci, S. (2005) 'Prognostic factors in multidisciplinary rehabilitation treatment in multiple sclerosis: an outcome study', *Multiple Sclerosis,* 11 (6) pp.719-724.

Great Britain. Department of Health (1999) *National Service Framework for Mental Health*. London: The Stationery Office.

Great Britain. Department of Health (2001) *National Service Framework for Older People*. London: The Stationery Office.

Great Britain. Department of Health (2001) *The expert patient: a new approach to chronic disease management for the 21st century*. London: The Stationery Office.

Great Britain. Department of Health (2005) *National Service Framework for Long-term Conditions*. London: The Stationery Office.

Gronwall, D. (1977) 'Paced auditory serial-addition task: a measure of recovery from concussion', *Perceptual and Motor Skills,* 44 (2) pp.367-373.

Gronwall, D. & Sampson, H. (1974) *The Psychological Effects of Concussion*. Oxford: Oxford University Press.

Grover, R. & Vriens, M. (2006) *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*. Thousand Oaks: Sage Publications.

Gullion, C. & Rush, A. (1998) 'Toward a generalizable model of symptoms in major depressive disorder', *Biological Psychiatry,* 44 (10) pp.959-972.

Haase, C., Tinnefeld, M., Daum, I., Ganz, R., Haupts, M. & Faustmann, P. (2004) 'Cognitive, but not mood dysfunction develops in multiple sclerosis during 7 years of follow-up', *European Neurology,* 52 (2) pp.92-95.

Hacker, D., Bol, L., Horgan, D. & Rakow, E. (2000) 'Test prediction and performance in a classroom context', *Journal of Educational Psychology,* 92 (1) pp.160-170.

Hacker, D., Bol, L. & Keener, M. (2008) 'Metacognition in education: a focus on calibration', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory*. New York: Psychology Press, pp. 429-455.

Haensch, C. & Jorg, J. (2006) 'Autonomic dysfunction in multiple sclerosis', *Journal of Neurology,* 253 (Suppl 1) pp.i3-i9.

Hagedorn, R. (2001) *Foundations for Practice in Occupational Therapy*. London: Churchill Livingstone.

Hanlon, R. (1996) 'Motor learning following unilateral stroke', *Archives of Physical Medicine and Rehabilitation,* 77 (8) pp.811-815.

Hannay, H., Howieson, D., Loring, D., Fischer, J. & Lezak, M. (2004) 'Neuropathology for neuropsychologists', in: Lezak, M., Howieson, D., Loring, D. Hannay, H. & Fisher, J. *Neuropsychological Assessment*. (2004). Oxford: Oxford University Press, pp.157-285.

Harris, C. & D'Eon, J. (2008) 'Psychometric properties of the Beck Depression Inventory - Second Edition (BDI-II) in individuals with chronic pain', *Pain*, 137 (3) pp.609-622.

Hart, J. (1965) 'Memory and the feeling-of-knowing experience', *Journal of Educational Psychology,* 56 (4) pp.208-216.

Haussleiter, I., Brune, M. & Juckel, G. (2009) 'Review: Psychopathology in multiple sclerosis: diagnosis, prevalence and treatment', *Therapeutic Advances in Neurological Disorders,* 2 (1) pp.13-29.

Heaton, R., Grant, I., & Matthews, C. (1986) 'Differences in Neuropsychological Test Performance associated with age, education and sex', in Grant, I & Adams, K (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders.* New York, Oxford University Press. pp. 100-120

Henry, J. & Beatty, W. (2006) 'Verbal fluency deficits in multiple sclerosis', *Neuropsychologia,* 44 (7) pp.1166-1174.

Herrmann, D. (1982) 'Know thy memory: The use of questionnaires to assess and study memory', *Psychological Bulletin,* 92 (2) pp.434-452.

Hertel, P. & Milan, S. (1994) 'Depressive deficits in recognition: Dissociation of recollection and familiarity', *Journal of Abnormal Psychology,* 103 (4) pp.736-742.

Hertzog, C. (2002) 'Metacognition in older adults: implications for application', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.169-196.

Hertzog, C. & Hultsch, D. (2000) 'Metacognition in adulthood and old age', in: Craik, F. & Salthouse, T. *The Handbook of Aging and Cognition,* 2nd edn. Mahwah: Lawrence Erlbaum, pp.417-466.

Hertzog, C., Hultsch, D. & Dixon, R. (1989) 'Evidence for the convergent validity of two self-report metamemory questionnaires', *Developmental Psychology,* 25 (5) pp.687-700.

Holmes, J., Madgwick, T. & Bates, D. (1995) 'The cost of multiple sclerosis', *British Journal of Medical Economics,* 8 pp.181-193.

Hoogervorst, E., Van Winsen, L., Eikelenboom, M., Kalkers, N., Uitdehaag, B. & Polman, C. (2001) 'Comparisons of patient self-report, neurologic examination, and functional impairment in MS', *Neurology,* 56 (7) pp.934-937.

Hox, J. (2002) *Multilevel Analysis: Techniques and Applications*. Mahwah: Lawrence Erlbaum Publishers.

Hoyle, R. (ed) (1995) *Structural Equation Modeling: Concepts, Issues, and Applications*. London: Sage Publications.

Hu, L. & Bentler, P. (1995) 'Evaluating model fit', in: Hoyle, R. (ed) (1995) *Structural Equation Modeling: Concepts, Issues, and Applications*. London: Sage Publications, pp.76-99.

Hull, J., Tedlie, J. & Lehn, D. (1995) 'Modeling the relation of personality variables to symptom complaints: The unique role of negative affectivity'. *Structural Equation Modelling: Concepts, Issues, and publications*, Thousand Oaks: Sage Publications, pp.217-235

Hunt, A. & Kingstone, A. (2004) 'Multisensory executive functioning', *Brain and Cognition*, 55 (2) pp.325-327.

Hutchinson, S. & Olmos, A. (1998) 'Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data', *Structural Equation Modeling: A Multidisciplinary Journal*, 5 (4) pp.344-364.

Hux, K., Reid, R. & Lugert, M. (1994) 'Self-instruction training following neurological injury', *Applied Cognitive Psychology*, 8 (3) pp.259-271.

Iacobucci, D. (2008) *Mediation Analysis*. Los Angelas: SAGE Publications Inc.

Isingrini, A. (2008) 'Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy', *Brain and Cognition*, 67 (1) pp76-87

Jacobs, L., Beck, R., Simon, J., Kinkel, R., Brownscheidle, C., Murray, T., Simonian, N., Slasor, P. & Sandrock, A. (2000) 'Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis', *The New England Journal of Medicine*, 343 (13) pp.898-904.

James, L., Mulaik, S. & Brett, J. (1982) *Causal Analysis: Assumptions, Models, And Data*. Beverly Hills: Sage Publications.

Janowsky, J., Shimamura, A. & Squire, L. (1989) 'Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients', *Psychobiology*, 17 (1) pp.3-11.

Jensen, A. & Rohwer Jr, W. (1966) 'The Stroop Color-Word Test: a review', *Acta Psychologica*, 25 (1) pp.36-93.

Johnson, M., Hayes, S., D Esposito, M. & Raye, C. (2000) 'Confabulation', *Handbook of Neuropsychology*, 2 pp.383-408.

Joiner, T., Schmidt, K. & Metalsky, G. (1994) ‘Low-end specificity of the beck depression inventory’, *Cognitive Therapy and Research*, 18 (1) pp.55-68.

Joint Formulary Committee (2009) *British National Formulary 57.* London: Pharmaceutical Press.

Joreskog, K. & Sorbom, D. (1989) *Lisrel 7: A Guide to the Program and its Application.* Chicago: SPSS.

Joreskog, K. & Sorbom, D. (1999) *Lisrel 8.30*. Chicago: Scientific Software International.

Julian, L., Merluzzi, N. & Mohr, D. (2007) 'The relationship among depression, subjective cognitive impairment, and neuropsychological performance in multiple sclerosis', *Multiple Sclerosis*, 13 (1) pp.81-86.

Kalmar, J., Gaudino, E., Moore, N., Halper, J. & Deluca, J. (2008) 'The relationship between cognitive deficits and everyday functional activities in multiple sclerosis', *Neuropsychology*, 22 (4) pp.442-449.

Kamil, M., Pearson, P., Barr, R. & Mosenthal, P. (2000) *Handbook of Reading Research Volume III*. Mahwah: Lawrence Erlbaum.

Kandel, E., Schwartz, J. & Jessell, T. (2000) *Principles of Neural Science*. New York: McGraw-Hill Medical.

Kappos, L., Polman, C., Freedman, M., Edan, G., Hartung, H., Miller, D., Montalban, X., Barkhof, F., Bauer, L. & Jakobs, P. (2006) 'Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes', *Neurology*, 67 (7) pp.1242-1249.

Karoly, P. (1993) 'Mechanisms of self-regulation: A systems view', *Annual Review of Psychology*, 44 pp.23-52.

Kelemen, W. & Weaver, C. (1997) 'Enhanced metamemory at delays: why do judgments of learning improve over time?', *Journal Of Experimental Psychology: Learning, Memory and Cognition*, 23 (6) pp.1394-1409.

Kelley, C. & Jacoby, L. (2000) 'Recollection and familiarity: process-dissociation', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory*. Oxford: Oxford University Press, pp.215-228.

Kelloway, E. (1998) *Using Lisrel for Structural Equation Modeling: A Researcher's Guide*. Thousand Oaks, California: Sage Publicaitons.

Kendall, P., Hollon, S., Beck, A., Hammen, C. & Ingram, R. (1987) 'Issues and recommendations regarding use of the Beck Depression Inventory', *Cognitive Therapy and Research*, 11 (3) pp.289-299.

Kennedy, M. (2001) 'Retrospective confidence judgements made by adults with traumatic brain injury: relative and absolute accuracy', *Brain Injury,* 15 (6) pp.469-487.

Kennedy, M. (2006) 'Managing memory and metamemory impairments in individuals with traumatic brain injury', *The ASHA Leader Online.* online resource: http://www.asha.org/Publications/leader/2006/061017/f061017b.htm [accessed February 22, 2010]

Kennedy, M., Carney, E. & Peters, S. (2003) 'Predictions of recall and study strategy decisions after diffuse brain injury', *Brain Injury,* 17 (12) pp.1043-1064.

Kennedy, M. & Yorkston, K. (2004) 'The effects of frontal injury on "on-line" self-monitoring during verbal learning by adults with diffuse brain injury', *Neuropsychological Rehabilitation*, 14 (4) pp.449-465.

Kenny, D. & McCoach, D. (2003) 'Effect of the number of variables on measures of fit in structural equation modeling', *Structural Equation Modeling*, 10 (3) pp.333-351.

Kesselring, J. (2004) 'Neurorehabilitation in multiple sclerosis - what is the evidence-base?', *Journal of Neurology*, 251 (Supp4) pp.25-29.

Kesselring, J. & Beer, S. (2005) 'Symptomatic therapy and neurorehabilitation in multiple sclerosis', *The Lancet Neurology*, 4 (10) pp.643-652.

Khan, F., Ng, L. & Turner-Stokes, L. (2009a) 'Effectiveness of vocational rehabilitation intervention on the return to work and employment of persons with multiple sclerosis', *Cochrane Database of Systematic Reviews.* 1:CD007256.

Khan, F., Pallant, J., Brand, C. & Kilpatrick, T. (2008a) 'Effectiveness of rehabilitation intervention in persons with multiple sclerosis: a randomised controlled trial', *Journal of Neurology, Neurosurgery & Psychiatry,* 79 (11) pp.1230-1235.

Khan, F., Turner-Stokes, L., Ng, L. & Kilpatrick, T. (2000) 'Multidisciplinary rehabilitation for adults with multiple sclerosis, the Cochrane Database of Systematic Reviews', in: *The Cochrane Library 2000 Issue 3*.

Khan, F., Turner-Stokes, L., Ng, L. & Kilpatrick, T. (2008b) 'Multidisciplinary rehabilitation for adults with multiple sclerosis', *Journal of Neurology, Neurosurgery & Psychiatry*, 79 (2) pp.114-121.

Khan, F., Turner-Stokes, L., Stevermuer, T. & Simmonds, F. (2009b) 'Multiple sclerosis rehabilitation outcomes: analysis of a national casemix data set from Australia', *Multiple Sclerosis,* 15 (7) pp.869-875.

Kim, K (2005) 'The relation among fit indices, power, and sample size in structural equation modeling' *Structural Equation Modeling.* 12 (3) pp. 368-390.

Kimball, D. & Metcalfe, J. (2003) 'Delaying judgments of learning affects memory, not metamemory', *Memory and Cognition,* 31 (6) pp.918-929.

Kit, K., Mateer, C. & Graves, R. (2007) 'The influence of memory beliefs in individuals with traumatic brain injury', *Rehabilitation Psychology,* 52 (1) pp.25-32.

Kline, R. (2005) *Principles and Practice of Structural Equation Modelling* (2005) 2nd edn. New York: The Guilford Press.

Kline, R. (2006) 'Reverse arrow dynamics. Formative measurement and feedback loops', in: Hancock, G. & Mueller, R. (eds) *Structural Equation Modeling: A Second Course.* Greenwich: Information Age Publishing, pp.43-68.

Kobelt, G., Berg, J., Lindgren, P., Fredrikson, S. & Jonsson, B. (2006) 'Costs and quality of life of patients with multiple sclerosis in Europe', *Journal of Neurology, Neurosurgery & Psychiatry,* 77 (8) pp.918-926.

Kobelt, G., Berg, J., Lindgren, P., Kerrigan, J., Russell, N. & Nixon, R. (2006) 'Costs and quality of life of multiple sclerosis in the United Kingdom', *The European Journal of Health Economics,* 7 (Suppl 2) pp.96-104.

Koren, D., Seidman, L., Goldsmith, M. & Harvey, P. (2006) 'Real-World Cognitive-and Metacognitive-Dysfunction in Schizophrenia: A New Approach for Measuring (and Remediating) More" Right Stuff", *Schizophrenia Bulletin,* 32 (2) pp.310-326.

Koriat, A. (1993) 'How do we know that we know? The accessibility model of the feeling of knowing', *Psychological Review,* 100 (4) pp.609-639.

Koriat, A. (1994) 'Memory's knowledge of its own knowledge: the accessibility account of the feeling of knowing', in: Metcalfe, J. & Shimamura, A. (eds) (1994) *Metacognition: Knowing about Knowing*. Cambridge MA: The MIT Press Cambridge, pp.115-135

Koriat, A. (1997) 'Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning', *Journal Of Experimental Psychology: General,* 126 (4) pp.349-370.

Koriat, A. (2000) 'Control processes in remembering', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory.* New York: Oxford University Press, pp.333-346.

Koriat, A. (2002) 'Metacognition research: an interim report', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.261-286.

Koriat, A. (2007) 'Metacognition and consciousness', in: Zelazo, P., Moscovitch, M. & Thompson, E. (eds) *The Cambridge Handbook of Consciousness,* pp.289-325.

Koriat, A. (2008) When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review,* 15 (5) pp.997-1001.

Koriat, A. & Bjork, R. (2005) 'llusions of competence in monitoring one's knowledge during study', *Journal Of Experimental Psychology: Learning Memory and Cognition,* 31(2) pp.187-194.

Koriat, A. & Bjork, R. (2006) 'Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test', *Memory and Cognition,* 34 (5) pp.959-972.

Koriat, A. & Goldsmith, M. (1994) 'Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment', *Journal Of Experimental Psychology: General,* 123 (3) pp.297-314.

Koriat, A. & Goldsmith, M. (1996) 'Monitoring and control processes in the strategic regulation of memory accuracy', *Psychological Review,* 103 (3) pp.490-517.

Koriat, A. & Goldsmith, M. (1998) 'The role of metacognitive processes in the regulation of memory performance', in: Mazzoni, G. & Nelson, T. (eds) *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes,* Mahwah: Lawrence Erlbaum, pp.97–118.

Koriat, A., Goldsmith, M. & Pansky, A. (2000) 'Toward a psychology of memory accuracy', *Annual Review of Psychology,* 51 pp.481-537.

Koriat, A. & Levy-Sadot, R. (1999) 'Processes underlying metacognitive judgments: information-based and experience-based monitoring of one's own knowledge', in: Chaiken, S. & Trope, Y. (eds) *Dual Process Theories In Social Psychology.* New York: The Guildford Press, pp.483-502.

Koriat, A. & Levy-Sadot, R. (2001) 'The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing', *Journal Of Experimental Psychology: Learning, Memory, and Cognition,* 27 (1) pp.34-53.

Koriat, A., Nussinson, R., Bless, H. & Shaked, N. (2008) 'Information-based and experience-based metacognitive judgments: evidence from subjective confidence', in: Dunlosky, J. & Bjock, R. (eds) *Handbook of Metamemory and Memory.* New York: Psychology Press, pp.117-135.

Koriat, A., Sheffer, L. & Ma'ayan, H. (2002) 'Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice', *Journal of Experimental Psychology: General,* 131 (2) pp.147-162.

Krakauer, J. (2006) 'Motor learning: its relevance to stroke recovery and neurorehabilitation', *Current Opinion in Neurology,* 19 (1) pp.84-90.

Kremenchutzky, M., Cottrell, D., Rice, G., Hader, W., Baskerville, J., Koopman, W. & Ebers, G. (1999) 'The natural history of multiple sclerosis: a geographically based study: 7. Progressive-relapsing and relapsing-progressive multiple sclerosis: a re-evaluation', *Brain,* 122 (10) pp.1941-1950.

Krinsky, R. & Nelson, T. (1985) 'The feeling of knowing for different types of retrieval failure', *Acta Psychologica,* 58 (2) pp.141-158.

Kruglanski, A. (1989) 'The psychology of being "right": The problem of accuracy in social perception and cognition', *Psychological Bulletin,* 106 (3) pp.395-409.

Kruglanski, A. (1999) 'Motivation, cognition, and reality: Three memos for the next generation of research', *Psychological Inquiry,* 10 (1) pp.54-58.

Krupp, L. & Christodoulou, C. (2001) 'Fatigue in multiple sclerosis', *Current Neurology and Neuroscience Reports,* 1 (3) pp.294-298.

Krupp, L., Christodoulou, C., Melville, P., Scherl, W., Macallister, W. & Elkins, L. (2004) 'Donepezil improved memory in multiple sclerosis in a randomized clinical trial', *Neurology,* 63 (9) pp.1579-1585.

Kurtzke, J. (1983) 'Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)', *Neurology,* 33 (11) pp.1444-1452.

Lachman, M. & Andreoletti, C. (2006) 'Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults', *Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 61 (2) pp.88-94.

Lachman, M., Andreoletti, C. & Pearman, A. (2006) 'Memory control beliefs: how are they related to age, strategy use and memory improvement?', *Social Cognition,* 24 (3) pp.359-385.

Lachman, M., Weaver, S., Bandura, M., Elliot, E. & Lewkowicz, C. (1992) 'Improving memory and control beliefs through cognitive restructuring and self-generated strategies', *The Journal of Gerontology,* 47 (5) pp.293-299.

Lad, F. (1984) 'The Calibration Question', *The British Journal for the Philosophy of Science,* 35 (3) pp.213-221.

Laming, D. (2009) 'Failure to recall', *Psychological Review,* 116 (1) pp.157-186.

Landro, N., Celius, E. & Sletvold, H. (2004) 'Depressive symptoms account for deficient information processing speed but not for impaired working memory in early phase multiple sclerosis (MS)', *Journal of the Neurological Sciences,* 217 (2) pp.211-216.

Landro, N., Sletvold, H. & Celius, E. (2000) 'Memory functioning and emotional changes in early phase multiple sclerosis', *Archives of Clinical Neuropsychology,* 15 (1) pp.37-46.

Lane, C. & Zelinski, E. (2003) 'Longitudinal hierarchical linear models of the memory functioning questionnaire', *Psychology and Aging,* 18 (1) pp.38-53.

Langdon, D. & Thompson, A. (1999) 'Multiple sclerosis: a preliminary study of selected variables affecting rehabilitation outcome', *Multiple Sclerosis,* 5 (2) pp.94-100.

LaRocca, N. (2000) 'Cognitive and emotional disorders', in: Burks, J. & Johnson, K. (eds) *Multiple Sclerosis: Diagnosis, Medical Management and Rehabilitation*. New York: Demos Medical Publishing, pp.405-424.

LaRocca, N., Kalb, R., Deluca, J. & Caruso, L. (2006) *Multiple Sclerosis: Understanding the Cognitive Challenges*. New York: Demos Medical Publishing.

Lau, H., Rogers, R., Haggard, P. & Passingham, R. (2004) 'Attention to intention', *Science,* 303 (5661) pp.1208-1210.

Lawton, G., Lundgren-Nilsson, A., Biering-Sorensen, F., Tesio, L., Slade, A., Penta, M., Grimby, G., Ring, H. & Tennant, A. (2006) 'Cross-cultural validity of FIM in spinal cord injury', *Spinal Cord,* 44 (12) pp.746-752.

Lengenfelder, J., Bryant, D., Diamond, B., Kalmar, J., Moore, N. & Deluca, J. (2006) 'Processing speed interacts with working memory efficiency in multiple sclerosis', *Archives of Clinical Neuropsychology,* 21 (3) pp.229-238.

Leonesio, R. (2008) 'Feeling-of-knowing accuracy and recollective experience', in: Dunlosky, J. & Bjock, R. (eds) *Handbook of Metamemory and Memory.* New York: Psychology Press, pp.195-210.

Leonesio, R. & Nelson, T. (1990) 'Do different metamemory judgments tap the same underlying aspects of memory?', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 16 (3) pp.464-470.

Lezak, M. (1983) *Neuropsychological Assessment 3rd edn.* New York: Oxford University Press.

Lezak, M., Howieson, D. & Loring, D. (2004) *Neuropsychological Assessment 4th edn.* New York: Oxford University Press.

Lincoln, N., Dent, A., Harding, J., Weyman, N., Nicholl, C., Blumhardt, L. & Playford, E. (2002) 'Evaluation of cognitive assessment and cognitive intervention for people with multiple sclerosis', *Journal of Neurology, Neurosurgery & Psychiatry,* 72 (1) pp.93-98.

Lineweaver, T. & Hertzog, C. (1998) 'Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs', *Aging, Neuropsychology and Cognition,* 5 (4) pp.264-296.

Lipinska, B. & Backman, L. (1997) 'Encoding-retrieval interactions in mild Alzheimer's disease: The role of access to categorical information', *Brain and Cognition,* 34 (2) pp.274-286.

Little, R. & Rubin, D. (1987) *Statistical analysis with Missing Data*. Hoboken: John Wiley & Sons.

Little, R. & Rubin, D. (1989) 'The analysis of social science data with missing values', *Sociological Methods & Research,* 18 (2-3) pp.292-326

Loftus, E. & Palmer, J. (1974) 'Reconstruction of automobile destruction: An example of the interaction between language and memory', *Journal of Verbal Learning and Verbal Behavior,* 13 (5) pp.585-589.

Lorig, K., Hurwicz, M., Sobel, D., Hobbs, M. & Ritter, P. (2005) 'A national dissemination of an evidence-based self-management program: a process evaluation study', *Patient Education and Counseling,* 59 (1) pp.69-79.

Lovera, J., Bagert, B., Smoot, K., Wild, K., Frank, R., Bogardus, K., Oken, B., Whitham, R. & Bourdette, D. (2006) 'Correlations of Perceived Deficits Questionnaire of multiple sclerosis Quality of Life Inventory with Beck Depression Inventory and neuropsychological tests', *Journal of Rehabilitation Research and Development,* 43 (1) pp.73-82.

Lowe, C. & Rabbitt, P. (1997) 'Cognitive models of aging and frontal lobe deficits', in: Rabbitt, P. (ed) *Methodology of Frontal and Executive Function.* Hove: Psychology Press, pp.39-59.

Lublin, F. & Reingold, S. (1996) 'Defining the clinical course of multiple sclerosis: results of an international survey', *Neurology,* 46 (4) pp.907-911.

Luria, A. (1980) *Higher Cortical Functions in Man. New York: Basic.*

Lycke, J., Tollesson, P. & Frisèn, L. (2001) 'Asymptomatic visual loss in multiple sclerosis', *Journal of Neurology,* 248 (12) pp1079-1086.

MacCallum, R. (1986) 'Specification searches in covariance structure modeling', *Psychological Bulletin,* 100 (1) pp.107-120.

MacCallum, R., Browne, M. & Sugawara, H. (1996) 'Power analysis and determination of sample size for covariance structure modeling', *Psychological Methods,* 1 (2) pp.130-149.

MacCallum, R., Roznowski, M. & Necowitz, L. (1992) 'Model modifications in covariance structure analysis: The problem of capitalization on chance', *Psychological Bulletin,* 111 (3) pp.490-504.

MacDonald, B., Cockerell, O., Sander, J. & Shorvon, S. (2000) 'The incidence and lifetime prevalence of neurological disorders in a prospective community-based study in the UK', *Brain,* 123 (4) pp.665-676.

MacKinnon, D., Krull, J. & Lockwood, C. (2000) Equivalence of the mediation, confounding and suppression effect', *Prevention Science,* 1 (4) pp.173-181.

MacKinnon, D., Lockwood, C., Hoffman, J., West, S. & Sheeis, V. (2002) 'A comparison of methods to test mediation and other intervening variable effects', *Psychological Methods,* 7 (1) pp.83-104.

MacKinnon, D., Lockwood, C. & Williams, J. (2004) 'Confidence intervals for the indirect effect: Distribution of the product and resampling methods', *Multivariate Behavioral Research,* 39 (1) pp.99-128.

MacKinnon, D. (2008) *Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum Associates.

MacNiven, J., Davis, C., Ho, M., Bradshaw, C., Szabadi, E. & Constantinescu, C. (2008) 'Stroop performance in multiple sclerosis: Information processing, selective attention, or executive functioning?', *Journal of the International Neuropsychological Society,* 14 (5) pp.805-814.

Maki, R. (2008) 'Privileged access for general knowledge and newly learned text material', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Memory and Metamemory*. New York: Psychology Press, pp.173-194

Maki, R. & McGuire, M. (2002) 'Metacognition for text: findings and implications for education', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.39-67.

Manchester, D., Priestley, N. & Jackson, H. (2004) 'The assessment of executive functions: coming out of the office', *Brain Injury,* 18 (11) pp.1067-1081.

Maor, Y., Olmer, L. & Mozes, B. (2001) 'The relation between objective and subjective impairment in cognitive function among multiple sclerosis patients - the role of depression', *Multiple Sclerosis*, 7 (2) pp.131-135.

Marcoulides, G. & Hershberger, S. (1997) *Multivariate Statistical Methods: A First Course*. Mahwah: Lawrence Erlbaum Associates.

Mardia, K. (1970) 'Measures of multivariate skewness and kurtosis with applications', *Biometrika*, 57 (3) pp.519-530.

Mardia, K. (1974) 'Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies', *Sankhya : The Indian Journal of Statistics,* 36 Series B (2) pp.115-128.

Mardia, K. (1980) 'Tests of univariate and multivariate normality', in: Krishnaiah, P. (ed) *The Handbook of Statistics Volume 1.* Amsterdam: North Holland, pp.279-320.

Markowitsch, H. (2003) 'Autonoetic consciousness', in: Kircher, T. & David, A. (eds) *The Self in Neuroscience and Psychiatry.* Cambridge: Cambridge University Press, pp.180-196.

Marois, R. & Ivanoff, J. (2005) 'Capacity limits of information processing in the brain', *Trends in Cognitive Sciences,* 9 (6) pp.296-305.

Marquie, J., Jourdan-Boddaert, L. & Huet, N. (2002) 'Do older adults underestimate their actual computer knowledge?', *Behaviour & Information Technology,* 21 (4) pp.273-280.

Marrie, R., Chelune, G., Miller, D. & Cohen, J. (2005) 'Subjective cognitive complaints relate to mild impairment of cognition in multiple sclerosis', *Multiple Sclerosis,* 11 (1) pp.69-75.

Marsh, H., Balla, J. & McDonald, R. (1988) 'Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size', *Psychological Bulletin*, 103 (3) pp.391-410.

Marsh, H. & Hau, K. (1999) 'Confirmatory factor analysis: strategies for small sample sizes', in: Hoyle, R. (ed) *Statistical Strategies For Small Sample Research.* Thousand Oaks: Sage Publications, pp.251-284.

Marsh, H., Hau. K, & Wen, Z. (2004). 'In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings', *Structural Equation Modeling*, 11 (3) pp.320-341.

Marsh, H., Hau, K. & Grayson, D. (2005) 'Goodness of fit evaluation in structural equation modeling', in: Maydeu-Olivares, A. & McArdle, J. (eds) *Contermpory Psychometrics. A festschrift to Roderick P. Mcdonald.* Mahwah: Lawrence Erlbaum Associates, pp.276-335.

Marsh, H., Wen, Z. & Hau, K. (2006) 'Structural equation models of latent interaction and quadratic effects', in: Hancock, G. & Mueller, R. (eds) *Structural Equation Modeling: A Second Course.* Greenwich: Information Age Publishing, pp.225-268.

Maruyama, G. (1998) *Basics of Structural Equation Modeling*. Thousand Oaks: Sage Publications.

Mazzoni, G. (2008) 'Metacognitive processes in creating false beliefs and false memories: the role of event plausibility', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Memory and Metamemory*. New York: Psychology Press, pp.315-331.

Mazzoni, G. & Kirsch, I. (2002) 'Autobiographical memories and beliefs: a preliminary metacognitive model', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.121-145.

Mazzoni, G. & Nelson, T. (1998) *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes*. Mahwah: Lawrence Erlbaum Associates.

McCabe, M., McKern, S. & McDonald, E. (2004) 'Coping and psychological adjustment among people with multiple sclerosis', *Journal of Psychosomatic Research,* 56 (3) pp.355-361.

McCarthy, G. (2008) ‘Prefrontal mechanisms for executive control over emotional distraction are altered in major depression’, *Psychiatry Research: Neuroimaging,* 163 (2) pp**.**143-155.

McCarthy, M., Beaumont, J., Thompson, R. & Peacock, S. (2005) 'Modality-specific aspects of sustained and divided attentional performance in multiple sclerosis', *Archives of Clinical Neuropsychology,* 20 (6) pp.705-718.

McCartney, K., Burchinal, M. & Bub, K. (2006) 'Best practices in quantitative methods for developmentalists', *Monographs of the Society for Research in Child Development,* 71 (3) pp1-145.

McDonald, W., Compston, A., Edan, G., Goodkin, D., Hartung, H., Lublin, F., Mcfarland, H., Paty, D., Polman, C. & Reingold, S. (2001) 'Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis', *Annals of Neurology,* 50 (1) pp.121-127.

McFarland, C., Ross, M. & Giltrow, M. (1992) 'Biased recollections in older adults: The role of implicit theories of aging', *Journal of Personality and Social Psychology,* 62 (5) pp.837-850.

McGuigan, C. & Hutchinson, M. (2006) 'Unrecognised symptoms of depression in a community based population with multiple sclerosis', *Journal of Neurology,* 253 (2) pp.219-223.

McHugh, J., Galvin, P. & Murphy, R. (2008) 'Retrospective comparison of the original and revised McDonald criteria in a general neurology practice in Ireland', *Multiple Sclerosis,* 14 (1) pp.81-85.

McIntosh, C. (2007) 'Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007)', *Personality and Individual Differences,* 42 (5) pp.859-867.

Meeter, M. & Nelson, T. (2003) 'Multiple study trials and judgments of learning', *Acta Psychologica,* 113 (2) pp.123-132.

Metcalfe, J. (2000) 'Metamemory: theory and data', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory*. Oxford: Oxford University Press, pp.197-211.

Metcalfe, J. (2008) 'Evolution of metacognition', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Memory and Metamemory.* New York: Psychology Press, pp.29-46.

Metcalfe, J. & Finn, B. (2008) 'Evidence that judgments of learning are causally related to study choice', *Psychonomic Bulletin and Review,* 15 (1) pp.174-179.

Metcalfe, J., Schwartz, B. & Joaquim, S. (1993) 'The cue-familiarity heuristic in metacognition', *Journal Of Experimental Psychology, Learning, Memory, and Cognition,* 19 (4) pp.851-864.

Micceri, T. (1989) 'The unicorn, the normal curve, and other improbable creatures', *Psychological Bulletin,* 105 (1) pp.156-166.

Middleton, L., Denney, D., Lynch, S. & Parmenter, B. (2006) 'The relationship between perceived and objective cognitive functioning in multiple sclerosis', *Archives of Clinical Neuropsychology,* 21 (5) pp.487-494.

Miller, A., Shapiro, S., Gershtein, R., Kinarty, A., Rawashdeh, H., Honigman, S. & Lahat, N. (1998) 'Treatment of multiple sclerosis with Copolymer-1 (Copaxone): implicating mechanisms of Th1 to Th2/Th3 immune-deviation', *Journal of Neuroimmunology,* 92 (1-2) pp.113-121.

Milner, B. (1971) 'Interhemispheric differences in the localization of psychological processes in man', *British Medical Bulletin,* 27 (3) pp.272-277.

Minden, S. (2000) 'Mood disorders in multiple sclerosis: diagnosis and treatment', *Journal of NeuroVirology,* 6 (Suppl 2) pp.160-167.

Minden, S. & Schiffer, R. (1990) 'Affective disorders in multiple sclerosis review and recommendations for clinical research', *Archives of Neurology,* 47 (1) pp.98-104.

Minden, S. & Schiffer, R. (1991) 'Depression and mood disorders in multiple sclerosis', *Cognitive and Behavioral Neurology,* 4 (1) pp.62-77.

Minden, S. & Schiffer, R. (1993) 'Depression and affective disorders in multiple sclerosis', in: Halbreich, U. (ed) *Multiple Sclerosis: A Neuropsychiatric Disorder*. Washington: American Psychiatric Association, pp.33-54.

Miner, A. & Reder, L. (1994) 'A new look at feeling of knowing: its metacognitive role in regulating question answering', in: Metcalfe, J. & Shimamura, A. (eds) *Metacognition: Knowing about Knowing,* Cambridge: The MIT Press, pp.47-70.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A. & Wager, T. (2000a) 'The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis', *Cognitive Psychology,* 41 (1) pp.49-100.

Miyake, A., Emerson, M. & Friedman, N. (2000b) 'Assessment of executive functions in clinical settings: problems and recommendations', *Seminars in Speech & Language,* 21 (2) pp.169-183.

Miyake, A. & Shah, P. (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge: Cambridge University Press.

Modirrousta, M. & Fellows, L. (2008) 'Medial prefrontal cortex plays a critical and selective role in 'feeling of knowing' meta-memory judgments', *Neuropsychologia,* 46 (12) pp.2958-2965.

Moffat, N (1992) 'Strategies of memory therapy', in: Wilson, B. & Moffat, N. (eds) *Clinical Management of Memory Problems*. 2nd edn. London: Chapman & Hall, pp.83-116.

Mohr, D. & Cox, D. (2001) Multiple sclerosis: 'Empirical literature for the clinical health psychologist', *Journal of Clinical Psychology,* 57 (4) pp.479-499.

Mohr, D., Goodkin, D., Likosky, W., Beutler, L., Gatto, N. & Langan, M. (1997) 'Identification of Beck Depression Inventory items related to multiple sclerosis', *Journal of Behavioural Medicine,* 20 (4) pp.407-414.

Mohr, D., Hart, S. & Goldberg, A. (2003) 'Effects of treatment for depression on fatigue in multiple sclerosis', *Psychosomatic Medicine,* 65 (4) pp.542-547.

Moscovitch, M. & Winocur, G. (2002) ‘The frontal cortex and working with memory’, in: Stuss, D. & Knight, R. (eds) *Principles of Frontal Lobe Function*. New York: Oxford University Press, pp.188-209.

Moulin, C. (2002) 'Sense and sensitivity: metacognition in Alzheimer's disease', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.197-223.

Moulin, C., Perfect, T. & Jones, R. (2000) 'Evidence for intact memory monitoring in Alzheimer's disease: metamemory sensitivity at encoding', *Neuropsychologia,* 38 (9) pp.1242-1250.

Moojiaart, A. (2003) 'Estimating the statistical power in small samples by empirical distributions' in: Yani, H., Okada, A., Shingemasu, K., kano, Y., & Meulman, J (Eds.) *New Developments in Psychometrics.* Tokyo, Springer Verlag pp. 149-156.

Mueller, R. (1996) *Basic Principles of Structural Equation Modeling: An introduction to LISREL and EQS*. New York: Springer-Verlag.

Mulligan, N. & Lozito, J. (2004) 'Self-generation and memory', in: Ross, B. (ed) *The Psychology of Learning and Motivation: Advances in Research and Theory Volume 45.* San Diego: Elsevier Academic Press, pp.175-214.

Multiple Sclerosis Society (1997). *Symptom Management Survey Multiple Sclerosis*. United Kingdom: unpublished report.

Muthèn, B., Kaplan, D. & Hollis, M. (1987) 'On structural equation modeling with data that are not missing completely at random', *Psychometrika,* 52 (3) pp.431-462.

Muthèn, L. & Muthèn, B. (2002) 'How to use a Monte Carlo study to decide on sample size and determine power' *Structural equation Modeling,* 9 (4) pp.599-620

Narens, L. (2002) 'The irony of measurement by subjective estimations', *Journal Of Mathematical Psychology,* 46 pp.769-788.

Narens, L., Nelson, T. & Scheck, P. (2008). 'Memory monitoring and the delayed JOL effect', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Memory and Metamemory.* New York: Psychology Press, pp.137-153.

National Collaborating Centre For Chronic Conditions (2004) *Multiple Sclerosis: National Clinical Guidelines for Management in Primary and Secondary Care.* Royal College of Physicians, London.

National Institute For Clinical Excellence (2002) Technology Appraisal Guidance No. 32 '*beta-Interferon and Glatiramer Acetate for the Treatment of Multiple Sclerosis'* National Institute for Clinical Excellence, London.

Naugle, R. & Chelune, G. (1990) 'Integrating neuropsychological and real-life data: a neuropsychological model for assessing everyday functioning' in: Tupper, D. & Cicerone, K. (eds) *The Neuropsychology of Everyday Life: Assessment and Basic Competencies.* Norwell: Kluwer Academic Publishers, pp.57-73.

Neisser, U. (1996) 'Remembering as doing', *Behavioral and Brain Sciences,* 19 (2) 203.

Nelson, L., & Willison, J. (1991) *'National Adult Reading Test (NART): Test Manual*' 2nd edition. Windsor, NFER Nelson.

Nelson, T. (1984) 'A comparison of current measures of the accuracy of feeling-of-knowing predictions', *Psychological Bulletin,* 95 (1) pp.109-133.

Nelson, T. (ed) (1992) *Metacognition: Core Readings*. Boston: Allyn & Bacon.

Nelson, T. (1996a) 'Consciousness and metacognition', *American Psychologist,* 51 (2) pp.102-116.

Nelson, T. (1996b) 'Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item comments on Schraw (1995)', *Applied Cognitive Psychology,* 10 (3) pp.257-260.

Nelson, T. & Dunlosky, J. (1991a) 'The delayed-JOL effect: When delaying your judgments of learning can improve the accuracy of your metacognitive monitoring', *Psychological Science,* 2 pp.267-270.

Nelson, T. & Dunlosky, J. (1991b) 'When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect", *Psychological Science,* 2 (4) pp.267-270.

Nelson, T. & Dunlosky, J. (1992) 'How shall we explain the delayed-judgment-of-learning effect?', *Psychological Science,* 3 (5) pp.317-318.

Nelson, T., Dunlosky, J., Graf, A. & Narens, L. (1994a) 'Utilization of metacognitive judgments in the allocation of study during multitrial learning', *Psychological Science,* 5 (4) pp.207-213.

Nelson, T., Gerler, D. & Narens, L. (1984) 'Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning', *Journal of Experimental Psychology: General,* 113 (2) pp.282-300.

Nelson, T., Leonesio, R., Landwehr, R. & Narens, L. (1986) 'A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12 (2) pp.279-287.

Nelson, T., Leonesio, R., Shimamura, A., Landwehr, R. & Narens, L. (1982) 'Overlearning and the feeling of knowing', *Journal of Experimental Psychology: Learing, Memory and Cognition,* 8 (4) pp.279-288.

Nelson, T. & Narens, L. (1980) 'A new technique for investigating the feeling of knowing', *Acta Psychologica,* 46 (1) pp.69-80.

Nelson, T. & Narens, L. (1990) 'Metamemory: A theoretical framework and new findings', *The Psychology of Learning and Motivation,* 26 pp.125-173.

Nelson, T. & Narens, L. (1994) 'Why investigate metacognition', in: Metcalfe, J. & Shimamura, A. (eds) *Metacognition: Knowing about Knowing.* Cambridge: The MIT Press, pp.1-25.

Nelson, T., Narens, L. & Dunlosky, J. (2004) 'A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM)', *Psychological Methods,* 9 (1) pp.53-69.

Nelson, T., Stuart, R., Howard, C. & Crowley, M. (1999) 'Metacognition and clinical psychology: A preliminary framework for research and practice', *Clinical Psychology & Psychotherapy*, 6 (2) pp.73-79.

Ness, J., Chabas, D., Sadovnick, A., Pohl, D., Banwell, B. & Weinstock-Guttman, B. (2007) 'Clinical features of children and adolescents with multiple sclerosis', *Neurology,* 68 16 (Suppl 2) pp.S37-S45.

Nevitt, J. & Hancock, G. (2001) 'Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling', *Structural Equation Modeling: A Multidisciplinary Journal,* 8 (3) pp.353-377.

Netemeyer, R., Bearden, W. & Sharma, S. (2003) *Scaling Procedures: Issues and Applications*. Thousand Oaks: Sage Publications.

Nietfeld, J., Enders, C. & Schraw, G. (2006) 'A Monte Carlo comparison of measures of relative and absolute monitoring accuracy', *Educational and Psychological Measurement,* 66 (2) pp.258-271.

Nocentini, U., Pasqualetti, P., Bonavita, S., Buccafusca, M., De Caro, M., Farina, D., Girlanda, P., Le Pira, F., Lugaresi, A. & Quattrone, A. (2006) 'Cognitive dysfunction in patients with relapsing-remitting multiple sclerosis', *Multiple Sclerosis*, 12 (1) pp.77-87.

Nolde, S., Johnson, M. & Raye, C. (1998) 'The role of prefrontal cortex during tests of episodic memory', *Trends in Cognitive Sciences*, 2 (10) pp.399-406.

Norman, D. & Shallice, T. (1986) 'Attention to action: willed and automatic control of behaviour', in: Davidson, R., Schwarts, G. & Shapiro. D. (eds) *Consciousness and self-regulation.* New York: Plenum Press, pp.1-18

Norris, G. & Tate, R. (2000) 'The Behavioural Assessment of the Dysexecutive Syndrome (BADS): ecological, concurrent and construct validity', *Neuropsychological Rehabilitation,* 10 (1) pp.33-45.

Noseworthy J., Lucchinetti. C., Rodriguez M., Weinshenker B. (2000) 'Multiple Sclerosis', *New England Journal of Medicine*, 343 (13) pp.938-952.

Nyenhuis, D., Rao, S., Zajecka, J. & Luchetta, T. (1995) 'Mood disturbance versus other symptoms of depression in multiple sclerosis', *Journal of the International Neuropsychological Society*, 1 (3) pp.291-296.

O'Hara, L., Cadbury, H., De Souza, L. & Ide, L. (2002) 'Evaluation of the effectiveness of professionally guided self-care for people with multiple sclerosis living in the community: a randomized controlled trial', *Clinical Rehabilitation*, 16 (2) pp.119-128.

O'Hara, L., De Souza, L. & Ide, L. (2000) 'A Delphi study of self-care in a community population of people with multiple sclerosis', *Clinical Rehabilitation*, 14 (1) pp.62-71.

Oberauer, K., Sub¸ H., Wilhelm, O. & Sander, N. (2007) 'Individual differences in working memory capacity and reasoning ability', in: Conway, A., Jarrold, C. & Kane, M. (eds) *Variation in Working Memory.* New York: Oxford University Press, pp.49-75.

O'Brien, A., Chiaravalloti, N., Goverover, Y. & Deluca, J. (2008) 'Evidenced-based cognitive rehabilitation for persons with multiple sclerosis: A review of the literature', *Archives of Physical Medicine and Rehabilitation,* 89 (4) pp.761-769.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M. & Yamadori, A. (2003) 'Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes', *Neuroimage*, 19 (4) pp.1369-1380.

Olivares, T., Nieto, A., Sanchez, M., Wollmann, T., Hernandez, M. & Barroso, J. (2005) 'Pattern of neuropsychological impairment in the early phase of relapsing-remitting multiple sclerosis', *Multiple Sclerosis*, 11 (2) pp.191-197.

Olmos, A. & Hutchinson, S. (1998) 'Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data', *Structural Equation Modeling*, 5 (4) pp.344-64.

Orton, S., Herrera, B., Yee, I., Valdar, W., Ramagopalan, S., Sadovnick, A. & Ebers, G. (2006) 'Sex ratio of multiple sclerosis in Canada: a longitudinal study', *The Lancet Neurology,* 5 (11) pp.932-936.

Osaka, N., Logie, R. & D'Esposito, M. (eds) (2007) *The Cognitive Neuroscience of Working Memory*. New York: Oxford University Press.

Oshinsky, R., Elfont, R. & Lublin, F. (1998) 'Incidence and clinical characteristics of the progressive relapsing form of MS', *Neurology,* 50: A209.

Osman, A., Downs, W., Barrios, F., Kopper, B., Gutierrez, P. & Chiros, C. (1997) 'Factor structure and psychometric characteristics of the Beck Depression Inventory-II', *Journal of Psychopathology and Behavioral Assessment*, 19 (4) pp.359-376.

Ownsworth, T. & McKenna, K. (2004) 'Investigation of factors related to employment outcome following traumatic brain injury: a critical review and conceptual model', *Disability & Rehabilitation*, 26 (13) pp.765-783.

Ownsworth, T. & Clare, L. (2006) 'The association between awareness deficits and rehabilitation outcome following acquired brain injury', *Clinical Psychology Review*, 26 (6) pp.783-795.

Ownsworth, T., Fleming, J., Desbois, J., Strong, J. & Kuipers, P. (2006) 'A metacognitive contextual intervention to enhance error awareness and functional outcome following traumatic brain injury: a single-case experimental design', *Journal of the International Neuropsychological Society*, 12 (1) pp.54-63.

Pannu, J. & Kaszniak, A. (2005) 'Metamemory experiments in neurological populations: a review', *Neuropsychology Review,* 15 (3) pp.105-130.

Pappas, B., Sunderland, T., Weingartner, H., Vitiello, B., Martinson, H. & Putnam, K. (1992) 'Alzheimer's disease and feeling-of-knowing for knowledge and episodic memory', *The Journal of Gerontology,* 47 (3) pp.159-164.

Parmenter, B., Weinstock-Guttman, B., Garg, N., Munschauer, F. & Benedict, R. (2007) 'Screening for cognitive impairment in multiple sclerosis using the Symbol Digit Modalities Test', *Multiple Sclerosis,* 13 (1) pp.52-57.

Patten, S. & Metz, L. (1997) 'Depression in multiple sclerosis', *Psychotherapy and Psychosomatics,* 66 (6) pp.286-292.

Patten, S. & Metz, L. (2001) 'Interferon {beta}-1a and depression in relapsing-remitting multiple sclerosis: an analysis of depression data from the PRISMS clinical trial', *Multiple Sclerosis,* 7 (4) pp.243-248.

Patten, S. & Metz, L. (2002) 'Interferon {beta} 1a and depression in secondary progressive MS: Data from the SPECTRIMS Trial', *Neurology,* 59 (5) pp.744-746.

Patti, F., Ciancio, M., Reggio, E., Lopes, R., Palermo, F., Cacopardo, M. & Reggio, A. (2002) 'The impact of outpatient rehabilitation on quality of life in multiple sclerosis', *Journal of Neurology,* 249 (8) pp.1027-1033.

Paty, D. (2000) 'Initial Symptoms', in: Burks, J. & Johnson, K. *Multiple Sclerosis: Diagnosis, Medical Management, and Rehabilitation*. New York: Demos Medical Publishing. pp.75-87.

Paul, R., Blanco, C., Hames, K. & Beatty, W. (1997) 'Autobiographical memory in multiple sclerosis', *Journal of the International Neuropsychological Society,* 3 (3) pp.246-251.

Pennington, B., Bennetto, L., Mcaleer, O. & Roberts, R. (1996) 'Executive functions and working memory: theoretical and measurement issues', in: Reid Lyon, G. & Krasnegor, N. (eds). *Attention, Memory, and Executive Function.* Baltimore: Paul H Brookes Publishing, pp.327-348.

Perfect, T. (2002) 'When does eyewitness confidence predict performance?', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.95-120.

Perfect, T. & Hollins, T. (1996) 'Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge', *Applied Cognitive Psychology,* 10 (5) pp.371-382.

Perfect, T. & Schwartz, B. (eds) (2002) 'Introduction: toward and applied metacognition', *Applied Metacognition.* Cambridge: Cambridge University Press, pp.1-14.

Perrotin, A., Belleville, S. & Isingrini, M. (2007) 'Metamemory monitoring in mild cognitive impairment: evidence of a less accurate episodic feeling-of-knowing', *Neuropsychologia,* 45 (12) pp.2811-2826.

Perrotin, A., Isingrini, M., Souchay, C., Clarys, D. & Taconnat, L. (2006) 'Episodic feeling-of-knowing accuracy and cued recall in the elderly: Evidence for double dissociation involving executive functioning and processing speed', *Acta Psychologica*, 122 (1) pp.58-73.

Peters, C. & Enders, C. (2002) 'A primer for the estimation of structural equation models in the presence of missing data: maximum likelihood algorithms', *Journal of Targeting, Measurement and Analysis for Marketing*, 11 (1) pp.81-95.

Petrides, M. (2000) 'Frontal lobes and memory', in: Cermak, L, (ed) *Handbook of Neuropsychology: Memory and its Disorders Volume2.* Amsterdam: Elsevier Science, pp.67-84.

Peyser, J., Rao, S., Larocca, N. & Kaplan, E. (1990) 'Guidelines for neuropsychological research in multiple sclerosis', *Archives of Neurology,* 47 (1) pp.94-97.

Phillips, L. & Stuifbergen, A. (2006) 'The influence of metamemory on the quality of life of persons with multiple sclerosis', *Journal of Neuroscience Nursing*, 38 (6) pp.428-434.

Pittock, S. & Rodriguez, M. (2008) 'Benign multiple sclerosis: a distinct clinical entity with therapeutic implications', in: Rodriguez, M. (ed) *Advances in Multiple Sclerosis and Experimental Demyelinating Disease.* Berlin: Springer, pp.1-17.

Pohl, P., Winstein, C. & Onla-Or, S. (1997) 'Sensory-motor control in the ipsilesional upper extremity after stroke', *NeuroRehabilitation*, 9 (1) pp.57-69.

Polman, C., Reingold, S., Edan, G., Filippi, M., Hartung, H., Kappos, L., Lublin, F., Metz, L., Mcfarland, H. & O'Connor, P. (2005) 'Diagnostic criteria for multiple sclerosis: 2005 revisions to the" McDonald Criteria", *Annals of Neurology,* 58 (6) pp.840-846.

Polman, C., Thompson, A., Murray, T., Bowling, A. & Noseworthy, J. (2006) *Multiple Sclerosis: The Guide to Treatment and Management*. 6th edn. New York: Demos Medical Publishing.

Ponsford, J. & Kinsella, G. (1992) 'Attentional deficits following closed-head injury', *Journal of Clinical and Experimental Neuropsychology,* 14 (5) pp.822-838.

Poser, C. & Brinar, V. (2001) 'Diagnostic criteria for multiple sclerosis', *Clinical Neurology and Neurosurgery,* 103 (1) pp.1-11.

Poskanzer, D., Sever, J., Sheridan, J. & Prenney, L. (1980) 'Multiple sclerosis in the Orkney and Shetland Islands. IV: Viral antibody titres and viral infections', *Journal of Epidemiology and Community Health*, 34 (4) pp.258-264.

Poskanzer, D., Walker, A., Yonkondy, J. & Sheridan, J. (1976) 'Studies in the epidemiology of multiple sclerosis in the Orkney and Shetland Islands', *Neurology*, 26 (6 PT 2) pp.14-17.

Preacher, K. & Leonardelli, G. (2006) Calculation for the Sobel test.[http://www.people.ku.edu/~preacher/sobel/sobel.htm] accessed 23 September 2009.

Prigatano, G. (1999) *Principles of Neuropsychological Rehabilitation*. New York: Oxford University Press.

Prigatano, G. & Kime, S. (2003) 'What do brain dysfunctional patients report following memory compensation training?', *NeuroRehabilitation,* 18 (1) pp.47-55.

Prins, F., Veenman, M. & Elshout, J. (2006) 'The impact of intellectual ability and metacognition on learning: New support for the threshold of problematicity theory.' *Learning and Instruction,* 16 (4) pp. 374-387.

Pucci, E., Branas, P., D'amico, R., Giuliani, G., Solari, A. & Taus, C. (2007) Amantadine for fatigue in multiple sclerosis. *Cochrane database of systematic reviews (Online)*,(1).

Quintana, S. & Maxwell, S. (1999) 'Implications of recent developments in structural equation modeling for counseling psychology', *The Counseling Psychologist,* 27 (4) pp.485-527.

Rabbitt, P. (1997) *Methodology of Frontal and Executive Function*. Hove: Psychology Press.

Rabbitt, P., Maylor, E., Mcinnes, L., Bent, N. & Moore, B. (1995) 'What goods can self-assessment questionnaires deliver for cognitive gerontology?' *Applied Cognitive Psychology,* 9 (7) pp.S127-S152.

Rabins, P. (1990) 'Euphoria in multiple sclerosis', in: Rao, S. *Neurobehavioral Aspects of Multiple Sclerosis.* New York: Oxford University Press, pp.180-185.

Ramirez-Lassepas, M., Tulloch, J., Quinones, M. & Snyder, B. (1992) 'Acute radicular pain as a presenting symptom in multiple sclerosis', *Archives of Neurology,* 49 (3) pp.255-258.

Rammohan, K., Rosenberg, J., Lynn, D., Blumenfeld, A., Pollak, C. & Nagaraja, H. (2002) 'Efficacy and safety of modafinil (Provigil (R)) for the treatment of fatigue in multiple sclerosis: a two centre phase 2 study', *Journal of Neurology, Neurosurgery, & Psychiatry,* 72 (2) pp.179-183.

Randolph, C., Tierney, M., Mohr, E. & Chase, T. (1998) 'The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity', *Journal of Clinical and Experimental Neuropsychology,* 20 (3) pp.310-319.

Randolph, J., Arnett, P. & Freske, P. (2004) 'Metamemory in multiple sclerosis: exploring affective and executive contributors', *Archives of Clinical Neuropsychology,* 19 (2) pp.259-279.

Randolph, J., Arnett, P. & Higginson, C. (2001) 'Metamemory and tested cognitive functioning in multiple sclerosis', *The Clinical Neuropsychologist,* 15 (3) pp.357-368.

Randolph, J., Arnett, P., Higginson, C. & Voss, W. (2000) 'Neurovegetative symptoms in multiple sclerosis relationship to depressed mood, fatigue, and physical disability', *Archives of Clinical Neuropsychology,* 15 (5) pp.387-398.

Rao, S. (1995) 'Neuropsychology of multiple sclerosis', *Current Opinion in Neurology,* 8 (3) pp.216-220.

Rao, S. (2004) 'Cognitive function in patients with multiple sclerosis: impairment and treatment', *International Journal of MS Care,* 1 pp.9-22.

Rao, S., Grafman, J., Digiulio, D., Mittenberg, W., Bernardin, L., Leo, G., Luchetta, T. & Unverzagt, F. (1993) 'Memory dysfunction in multiple sclerosis: Its relation to working memory, semantic encoding, and implicit learning', *Neuropsychology,* 7 (3) pp.364-364.

Rao, S., Hammeke, T. & Speech, T. (1987) 'Wisconsin Card Sorting Test performance in relapsing-remitting and chronic-progressive multiple sclerosis', *Journal Consulting Clinical Psychology,* 55 (2) pp.263-265.

Rao, S., Leo, G., Bernardin, L. & Unverzagt, F. (1991a) 'Cognitive dysfunction in multiple sclerosis: I. Frequency, patterns, and prediction', *Neurology,* 41 (5) pp.685-691.

Rao, S., Leo, G., Ellington, L., Nauertz, T., Bernardin, L. & Unverzagt, F. (1991b) 'Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning', *Neurology,* 41 (5) pp.692-696.

Rao, S., Leo, G. & St Aubin-Faubert, P. (1989) 'On the nature of memory disturbance in multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology,* 11 (5) pp.699-712.

Raskin, S., & Mateer, C. (1994) 'Rehabilitation of cognitive imapirments', in: Good, D., & Couch, J. (eds) *Handbook of Neurorehabilitation.* New York: Marcel Dekker, pp.243-259.

Raykov, T. & Marcoulides, G. (2006) *A First Course in Structural Equation Modeling*. 2nd edn. Mahwah: Lawrence Erlbaum.

Raymond, P., Stern, R., Authelet, A. & Penney, D. (1987) 'A comparison of California Verbal Learning Test performance among patients with multiple sclerosis, right hemisphere vascular lesions, and normal controls', *International Neuropsychological Society meeting*, Washington DC.

Reder, L. & Ritter, F. (1992) 'What determines initial feeling of knowing? Familiarity with question terms, not with the answer', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18 (3) pp.435-451.

Reder, L. & Schunn, C. (1996) 'Metacognition does not imply awareness: strategy choice is governed by implicit learning and memory', in: Reder, L. (ed) *Implicit Memory and Metacognition.* Mahwah: Lawrence Erlbaum, pp.45-76.

Reulen, J., Sanders, E. & Hogenhuis, L. (1983) 'Eye movement disorders in multiple sclerosis and optic neuritis', *Brain*, 106 (1) pp.121-140.

Richards, R., Sampson, F., Beard, S. & Tappenden, P. (2002) 'A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models', *Health Technology Assessment*, 6 (10) pp.1-73.

Richardson, J. & Chan, R. (1995) 'The constituent structure of subjective memory questionnaires: evidence from multiple sclerosis', *Memory*, 3 (2) pp.187-200.

Richardson, J., Engle, R., Hasher, L., Logie, R., Stoltzfus, E. & Zacks, R. (1996) *Working Memory and Human Cognition*, New York: Oxford University Press.

Robertson, N., Fraser, M., Deans, J., Clayton, D., Walker, N. & Compston, D. (1996) 'Age-adjusted recurrence risks for relatives of patients with multiple sclerosis', *Brain*, 119 (2) pp.449-455.

Roediger, H., Bergman, E. & Meade, M. (2000) 'Repeated reproduction from memory', in: Saito, A. (ed) *Bartlett, Culture and Cognition.* Guildford: Psychology Press, pp.115-134.

Roediger, H. & McDermott, K. (2000) 'Distortions of memory', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory*. Oxford: Oxford University Press, pp.149-162.

Rogers, A., Kennedy, A., Bower, P., Gardner, C., Gately, C., Lee, V., Reeves, D. & Richardson, G. (2008) 'The United Kingdom Expert Patients Programme: results and implications from a national evaluation', *Medical Journal of Australia*, 189 (Suppl 10) pp.S21-S24.

Rosti, E., Hamalainen, P., Koivisto, K. & Hokkanen, L. (2006) 'The PASAT performance among patients with multiple sclerosis: analyses of responding patterns using different scoring methods', *Multiple Sclerosis*, 12 (5) pp.586-593.

Roth, P. (1994) 'Missing data: a conceptual review for applied psychologists', *Personnel Psychology*, 47 (3) pp.537-560.

Rotstein, Z., Hazan, R., Barak, Y. & Achiron, A. (2006) 'Perspectives in multiple sclerosis health care: Special focus on the costs of multiple sclerosis', *Autoimmunity Reviews*, 5 (8) pp.511-516.

Rowland, S., Lam, C. & Leahy, B. (2005) 'Use of the Beck Depression Inventory-II (BDI-II) with persons with traumatic brain injury: analysis of factorial structure', *Brain Injury*, 19 (2) pp.77-83.

Royall, D., Lauterbach, E., Cummings, J., Reeve, A., Rummans, T., Kaufer, D., Lafrance Jr, W. & Coffey, C. (2002) 'Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association', *Journal of Neuropsychiatry and Clinical Neurosciences*, 14 (4) pp.377-405.

Rubin, D. (1976) 'Inference and missing data', *Biometrika*, 63 (3) pp.581-592.

Rubin, D. (1996) 'Multiple imputation after 18+ years', *Journal of the American Statistical Association*, 91 (434) pp.473-489.

Russo, F. (2008) *Causality and Causal Modelling in the Social Sciences: Measuring Variations*. Berlin: Springer Verlag.

Sacks, T., Clark, C., Pols, R. & Geffen, L. (1991) 'Comparability and stability of performance of six alternate forms of the Dodrill-Stroop Colour-Word Test', *Clinical Neuropsychologist*, 5 (3) pp.220-225.

Sadovnick, A. & Ebers, G. (1993) 'Epidemiology of multiple sclerosis: a critical overview', *The Canadian Journal of Neurological Sciences*, 20 (1) pp.17-29.

Saklofske, D., Hildebrand, D. & Gorsuch, R. (2000) 'Replication of the factor structure of the Wechsler Adult Intelligence Scale - Third Edition with a Canadian sample. *Psychological Assessment*, 12 (4) pp.436-439.

Salokangas, R., Vaahtera, K., Pacriev, S., Solhman, B., & Lehtinen, V., (2002). 'Gender differences in depressive symptoms. An artefact caused by measurement instruments? *Journal of Affective Disorders.* 68 pp. 215-220.

Salthouse, T. (1991) 'Mediation of adult age differences in cognition by reductions in working memory and speed of processing', *Psychological Science*, 2 (3) pp.179-183.

Salthouse, T. (1991) ‘Expertise as the circumvention of human processing limitations’, in: Ericsson, K. & Smith, J. (eds) *Toward a General Theory of Expertise: Prospects and Limits*. Cambridge: Cambridge University Press, pp.286-300.

Salthouse, T. (1996) 'The processing-speed theory of adult age differences in cognition', *Psychological Review,* 103 (3) pp.403-428.

Salthouse, T. (2010) 'Does the meaning of Neurocognitive change change with age?' *Neuropsychology,* 24 (2) pp. 273-278.

Sanfilipo, M., Benedict, R., Weinstock-Guttman, B. & Bakshi, R. (2006) 'Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis', *Neurology,* 66 (5) pp.685-692.

Satorra, A. & Bentler, P. (1994) 'Corrections to test statistics and standard errors in covariance structure analysis', in: Von Eye, A. & Clogg, C. (eds) *Latent Variables Analysis: Applications for Developmental Research.* Thousand Oaks: Sage Publications, pp.399-419.

Sayao, A., Devonshire, V. & Tremlett, H. (2007) 'Longitudinal follow-up of "benign" multiple sclerosis at 20 years', *Neurology*, 68 (7) pp.496-500.

Sbordone, R., Saul, R. & Purisch, A. (2007) *Neuropsychology for Psychologists, Health Care Professionals, and Attorneys*. 3rd edn. Boca Raton: CRC Press.

Scarrabelotti, M. & Carroll, M. (1998) 'Awareness of remembering achieved through automatic and conscious processes in multiple sclerosis', *Brain and Cognition*, 38 (2) pp.183-201.

Scarrabelotti, M. & Carroll, M. (1999) 'Memory dissociation and metamemory in multiple sclerosis', *Neuropsychologia*, 37 (12) pp.1335-1350.

Schafer, J. (1997) *Analysis of Incomplete Multivariate Data*. Boca Ratan: Chapman & Hall/CRC.

Schafer, J. & Graham, J. (2002) 'Missing data: our view of the state of the art', *Psychological Methods*, 7 (2) pp.147-177.

Schapiro, R., (2006) 'Managing the symptoms of multiple sclerosis', in: Cook, S. (ed) *Handbook of Multiple Sclerosis*. 4th edn. New York: Taylor & Francis Group, pp.271-278.

Schapiro, R. (2007) *Managing the Symptoms of Multiple Sclerosis*. 5th edn. New York: Demos Medical Publishing Inc.

Schear, J. & Sato, S. (1989) 'Effects of visual acuity and visual motor speed and dexterity on cognitive test performance', *Archives of Clinical Neuropsychology,* 4 (1) pp.25-32.

Schinka, J., Velicer, W. & Weiner, I. (eds) (2003) *Handbook Of Psychology:Volume 2 Research Methods in Psychology*. Hoboken: John Wiley & Sons.

Schneider, W. (1998) 'The development of procedural metamemory in childhood and adolescence', in: Mazzoni, G & Nelson, T. (eds) *Metacognition and Cognitive Neuropsychology*. *Monitoring and Control Processes.* Mahwah: Lawrence Erlbaum Associates.pp.1-21.

Schneider, W. & Lockl, K. (2008) 'Procedural metacognition in children: evidence for developmental trends', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory.* Hove: Psychology Press, pp.391-409.

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

Schnyer, D., Verfaellie, M., Alexander, M., Lafleche, G., Nicholls, L. & Kaszniak, A. (2004) 'A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex', *Neuropsychologia,* 42 (7) pp.957-966.

Schraw, G. (1995) 'Measures of feeling-of-knowing accuracy: a new look at an old problem', *Applied Cognitive Psychology,* 9 (4) pp.321-332.

Schraw, G. (2009) 'A conceptual analysis of five measures of metacognitive monitoring', *Metacognition and Learning,* 4 (1) pp.33-45.

Schreurs, K., De Ridder, D. & Bensing, J. (2002) 'Fatigue in multiple sclerosis: Reciprocal relationships with physical disabilities and depression', *Journal of Psychosomatic Research,* 53 (3) pp.775-781.

Schubert, D. & Foliart, R. (1993) 'Increased depression in multiple sclerosis patients: a meta-analysis', *Psychosomatics,* 34 (2) pp.124-130.

Schumacker, R. & Lomax, R. (2004) *A Beginner's Guide to Structural Equation Modeling.* Mahwah: Lawrence Erlbaum Associates.

Schwartz, B. & Bacon, E. (2008) 'Metacognitive neuroscience', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory.* Hove: Psychology Press, pp.355-371.

Schwartz, C. & Fox, B. (1995) 'Who says yes? identifying selection biases in a psychosocial intervention study of multiple sclerosis', *Social Science & Medicine,* 40 (3) pp.359-370.

Schwartz, B. & Metcalfe, J. (1992) 'Cue familiarity but not target retrievability enhances feeling-of-knowing judgments', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 18 (5) pp.1074-1083.

Schwartz, B. & Metcalfe, J. (1994) 'Methodological problems and pitfalls in the study of human metacognition', in: Metcalfe, J. & Shimamura, A. (eds) *Metacognition: Knowing about Knowing.* Cambridge: The MIT Press, pp.93-113.

Seignourel, P., Green, C. & Schmitz, J. (2007) 'Factor structure and diagnostic efficiency of the BDI-II in treatment-seeking substance users', *Drug and Alcohol Dependence,* 93 (3) pp.271-278.

Sepulcre, J., Vanotti, S., Hernandez, R., Sandoval, G., Caceres, F., Garcea, O. & Villoslada, P. (2006) 'Cognitive impairment in patients with multiple sclerosis using the Brief Repeatable Battery-Neuropsychology test', *Multiple Sclerosis,* 12 (2) pp.187-195.

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

Shallice, T. & Burgess, P. (1991) 'Higher-order cognitive impairments and frontal lobe lesions in man', in: Levin, H., Eisenberg, H. & Benton, A. (eds) *Frontal Lobe Function and Dysfunction*. New York: Oxford University Press, pp.125-138.

Shallice, T. & Burgess, P. (1993) 'Supervisory control of action and thought selection', in: Baddeley, A. & Weiskrantz, L. *Attention: Selection, Awareness, and Control: A Tribute to Donald Broadbent,* New York: Oxford University Press, pp.171-187.

Shallice, T., Burgess, P. & Robertson, I. (1996) 'The domain of supervisory processes and temporal organization of behaviour', in: *Philosophical Transactions: Biological Sciences,* 351(1346) pp.1405-1412.

Schapiro, R. & Schneider, D. (1999) 'Fatigue' in: van den Noort, S. & Holland, N. *Multiple Sclerosis in Clinical Practice,* New York: Demos Medical Publishing Inc, pp.35-40.

Shepherd, D. & Summers, A. (1996) 'Prevalence of multiple sclerosis in Rochdale', *Journal of Neurology, Neurosurgery & Psychiatry,* 61 (4) pp.415-417.

Sheridan, L., Fitzgerald, H., Adams, K., Nigg, J., Martel, M., Puttler, L., Wong, M. & Zucker, R. (2006) 'Normative Symbol Digit Modalities Test performance in a community-based sample', *Archives of Clinical Neuropsychology,* 21 (1) pp.23-28.

Sherman, E., Strauss, E., Spellacy, F. & Hunter, M. (1995) Construct validity of WAIS-R factors: Neuropsychological test correlates in adults referred for evaluation of possible head injury. *Psychological Assessment,* 7 (4) pp.440-444.

Shimamura, A. (1996) 'Memory and the Prefrontal Cortex', *Annals of the New York Academy of Sciences,* 769 (1) pp.151-160.

Shimamura, A. (2000) 'Toward a Cognitive Neuroscience of Metacognition', *Consciousness and Cognition,* 9 (2) pp.313-323.

Shimamura, A. (2008) 'A neurocognitive approach to metacognitive monitoring and control', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory*. Hove: Psychology Press, pp.373-390.

Shimamura, A. & Squire, L. (1986) 'Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 12 (3) pp.452-460.

Siegert, R. & Abernethy, D. (2005) 'Depression in multiple sclerosis: a review', *British Medical Journal,* 76 (4) pp.469-475.

Smith, A. (1982) *A Symbol-Digit Modalities Test Manual.* Los Angeles: Western Psychological Services.

Smith, C., Samkoff, L. & Scheinberg, L. (1993) 'Clinical features, assessment, and differential diagnosis of multiple sclerosis', in: Halbreich, U. *Multiple Sclerosis: A Neuropsychiatric Disorder.* Washington: American Psychiatric Publications, pp.1-31

Smith, M. & Arnett, P. (2005) 'Factors related to employment status changes in individuals with multiple sclerosis', *Multiple Sclerosis,* 11 (5) pp.602-609.

Sniehotta, F., Scholz, U. & Schwarzer, R. (2005) 'Bridging the intention-behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise', *Psychology & Health,* 20 (2) pp.143-160.

Sobel, M. (1982) 'Asymptotic confidence intervals for indirect effects in structural equation models', *Sociological Methodology,* 13 pp.290-312.

Sobel, M. (1988) 'Direct and indirect effects in linear structural equation models', *Sociological Methods & Research,* 16 (1) pp.155-176.

Sollom, A. & Kneebone, I. (2007) 'Treatment of depression in people who have multiple sclerosis', Multiple *Sclerosis,* 13 (5) pp.632-635.

Son, L. & Kornell, N. (2008) 'Research on the allocation of study time: key studies from 1890 to the present (and beyond)', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory.* Hove: Psychology Press, pp.333-351.

Son, L. & Schwartz, B. (2002) 'The relation between metacognitive monitoring and control', in: Perfect, T. & Schwartz, B. (eds) *Applied Metacognition.* Cambridge: Cambridge University Press, pp.15-38.

Souchay, C. (2007) 'Metamemory in Alzheimer's disease', *Cortex,* 43 (7) pp.987-1003.

Souchay, C., Isingrini, M., Pillon, B. & Gil, R. (2003) 'Metamemory accuracy in Alzheimer's Disease and frontoltemporal lobe dementia', *Neurocase,* 9 (6) pp.482-492.

Souchay, C. & Isingrini, M. (2004) 'Age-related differences in the relation between monitoring and control of learning', *Experimental Aging Research,* 30 (2) pp.179-193.

Souchay, C., Isingrini, M. & Espagnet, L. (2000) 'Aging, episodic memory feeling-of-knowing, and frontal functioning', *Neuropsychology,* 14 (2) pp.299-309.

Souchay, C., Isingrini, M. & Gil, R. (2002) 'Alzheimer’s disease and feeling-of-knowing in episodic memory', *Neuropsychologia,* 40 (13) pp.2386-2396.

Souchay, C., Isingrini, M. & Gil, R. (2006) 'Metamemory monitoring and Parkinson's disease', *Journal of Clinical and Experimental Neuropsychology,* 28 (4) pp.618-630.

Souchay, C., Moulin, C., Clarys, D., Taconnat, L. & Isingrini, M. (2007) 'Diminished episodic memory awareness in older adults: evidence from feeling-of-knowing and recollection', *Consciousness and Cognition,* 16 (4) pp.769-784.

Spellman, B. & Bjork, R. (1992) 'When predictions create reality: Judgments of learning may alter what they are intended to assess', *Psychological Science,* 3 (5) pp.315-316.

Spellman, B., Bloomfield, A. & Bjork, R. (2008) 'Measuring memory and metamemory: theoretical and statistical problems with assessing learning (in general) and using gamma (in particular) to do so', in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory.* Hove: Psychology Press, pp.95-114.

Spinella, M. (2005) 'Self-rated executive function: Development of the executive function index', *International Journal of Neuroscience,* 115 (5) pp.649-667.

Spreen, O. & Strauss, E. (1998) *A Compendium of Neuropsychological Tests*: *Administration, Norms, and Commentary.* 2nd edn. New York: Oxford University Press.

SPSS Inc. (2007) *SPSS 16.0 for Macintosh.* Chicago: SPSS Inc.

Stankoff, B., Waubant, E., Confavreux, C., Edan, G., Debouverie, M., Rumbach, L., Moreau, T., Pelletier, J., Lubetzki, C. & Clanet, M. (2005) 'Modafinil for fatigue in MS: a randomized placebo-controlled double-blind study', *Neurology,* 64 (7) pp.1139-1143.

Stankov, L (2000) 'Complexity, Metacognition & Fluid Intelligence' *Intelligence.* 28 (2) pp. 121-143.

Starck, M., Albrecht, H., Pollmann, W., Straube, A. & Dieterich, M. (1996) 'Drug therapy for acquired pendular nystagmus in multiple sclerosis', *Journal of Neurology,* 244 (1) pp.9-16.

Steel, R. & Torrie, J. (1960) *Principles and Procedures of Statistics,* New York: McGraw-Hill.

Steer, R., Ball, R., Ranieri, W. & Beck, A. (1999) 'Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients', *Journal Of Clinical Psychology,* 55 (1) pp.117-128.

Steiger, J. (1990) 'Structural model evaluation and modification: an interval estimation approach', *Multivariate Behavioral Research,* 25 (2) pp.173-180.

Steiger, J. (2000) 'Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser', *Structural Equation Modeling*, 7 (2) pp.149-162.

Steiger, K. (2007) 'Information processing in multiple sclerosis: Accuracy versus speed',

Steultjens, E., Dekker, J., Bouter, L., Cardol, M., Van De Nes, J. & Van Den Ende, C. (2003) Occupational Therapy for multiple sclerosis. *The Cochrane Database of Systematic Reviews*, 31465-1858.

Stevens, J. (2002) *Applied Multivariate Statistics for the Social Sciences*. 4th edn. Mahwah: Lawrence Erlbaum.

Stewart, S. (2005) 'The effects of benzodiazepines on cognition', *Journal of Clinical Psychiatry*, 66 (suppl 2) pp.9-13.

Stokx, L. & Gaillard, A. (1986) 'Task and driving performance of patients with a severe concussion of the brain', *Journal of Clinical and Experimental Neuropsychology*, 8 (4) pp.421-436.

Storch, E., Roberti, J. & Roth, D. (2004) 'Factor structure, convergent validity, and internal consistency of the Beck Depression Inventory - second edition, in a sample of college students', *Depression and Anxiety*, 19 (3) pp.187-189.

Strauss, E., Sherman, E. & Spreen, O. (2006) *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd edn. New York: Oxford University Press.

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

Strunk, D. & Adler, A. (2009) 'Cognitive biases in three prediction tasks: A test of the cognitive model of depression', *Behaviour Research and Therapy*, 47 (1) pp.34-40.

Stuifbergen, A., Becker, H., Blozis, S., Timmerman, G. & Kullberg, V. (2003) 'A randomized clinical trial of a wellness intervention for women with multiple sclerosis', *Archives of Physical Medicine and Rehabilitation*, 84 (4) pp.467-476.

Stuss, D., Alexander, M., Palumbo, C., Buckle, L., Sayer, L. & Pogue, J. (1994) 'Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks', *Neuropsychology*, 8 (3) pp.355-373.

Stuss, D., Floden, D., Alexander, M., Levine, B. & Katz, D. (2001) 'Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location', *Neuropsychologia*, 39 (8) pp.771-786.

Stuss, D. & Knight, R. (eds) (2002) *Principles of Frontal Lobe Function*. New York: Oxford University Press.

Surguladze, S., Keedwell, P. & Phillips, M. (2003) 'Neural systems underlying affective disorders', *Advances in Psychiatric Treatment*, 9 (6) pp.446-455.

Surridge, D. (1969) 'An investigation into some psychiatric aspects of multiple sclerosis', *British Journal of Psychiatry*, 115 (524) pp.749-764.

Swirsky-Sacchetti, T., Field, H., Mitchell, D., Seward, J., Lublin, F., Knobler, R. & Gonzalez, C. (1992) 'The sensitivity of the Mini-Mental State Exam in the white matter dementia of multiple sclerosis', *Journal of Clinical Psychology*, 48 (6) pp.779-786.

Talley, C. (2005) 'The Emergence of Multiple Sclerosis', *Perspectives in Biology and Medicine*, 48 (3) pp.383-395.

The Psychological Corporation (2002) *WAIS-III Administration and Scoring Manual*. San Antonio: The Psychological Corporation.

Thiede, K., Anderson, M. & Therriault, D. (2003) 'Accuracy of metacognitive monitoring affects learning of texts', *Journal of Educational Psychology*, 95 (1) pp.66-73.

Thomas, P., Thomas, S., Hillier, C., Galvin, K. & Baker, R. (2006) Psychological interventions for multiple sclerosis. *Cochrane database of systematic reviews (Online)*,(1).

Thombs, B., Ziegelstein, R., Beck, C. & Pilote, L. (2008) A general factor model for the Beck Depression Inventory-II: Validation in a sample of patients hospitalized with acute myocardial infarction. *Journal of Psychosomatic Research,* 65,(2) 115-121.

Thompson, A. (2005) 'Neurorehabilitation in multiple sclerosis: foundations, facts and fiction', *Current Opinion in Neurology,* 18 (3) pp.267-271.

Thornton, A. & DeFreitas, V. (2009) 'The neuropsychology of multiple sclerosis', in: Grant, I. & Adams, K. *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders.* 3rd edn. New York: Oxford University Press.pp.280-305.

Thornton, A. & Raz, N. (1997) 'Memory impairment in multiple sclerosis: A quantitative review', *Neuropsychology,* 11 (3) pp.357-366.

Toglia, J. (1991) 'Generalization of treatment: a multicontext approach to cognitive perceptual impairment in adults with brain injury', *The American Journal of Occupational Therapy,* 45 (6) pp.505-516.

Toglia, J. (1998a) 'A dynamic interactional model to cognitive rehabilitation', in: Katz, N. (ed) *Cognition and Occupation in Rehabilitation: Cognitive Models for Intervention in Occupational Therapy.* Bethesda: American Occupational Therapy Association, pp.5-50.

Toglia, J. (1998b) 'The multicontext treatment approach', in: Willard. H. & Spackman, C. (eds) *Occupational Therapy.* 9th edn. Philadelphia: Lippincott Williams & Wilkins, pp.557-559.

Toglia, J. & Kirk, U. (2000) 'Understanding awareness deficits following brain injury', *NeuroRehabilitation,* 15 (1) pp.57-70.

Tomassini, V. & Pozzilli, C. (2009) 'Sex hormones, brain damage and clinical course of multiple sclerosis', *Journal of the Neurological Sciences*, 286 (1-2) pp.35-39.

Tullman, M., Oshinsky, R., Lublin, F. & Cutter, G. (2004) 'Clinical characteristics of progressive relapsing multiple sclerosis', *Multiple Sclerosis,* 10 (4) pp.451-454.

Tulving, E., Kapur, S., Craik, F., Moscovitch, M. & Houle, S. (1994) 'Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings', *Proceedings of the National Academy of Sciences,* 91 (6) pp.2016-2020

Tulving, E. & Madigan, S. (1970) 'Memory and verbal learning', *Annual Review of Psychology,* 21 pp.437-484.

Turley-Ames, K. & Whitfield, M. (2003) 'Strategy training and working memory task performance', *Journal of Memory and Language,* 49 (4) pp.446-468.

Ullman, J. & Bentler, P. (2004) 'Structural equation modeling', in: Hardy, M. & Bryman, A. (eds) *Handbook Of Data Analysis.* Thousand Oaks: Sage Publication, pp431-458.

Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y. & Fontaine, J. (2008) 'The factor structure of the Beck Depression Inventory-II: an evaluation.' *Assessment,* 15**,** (2) pp.177-187

Vakil, E. & Blachstein, H. (1993) 'Rey auditory-verbal learning test: structure analysis', *Journal of Clinical Psychology,* 49 (6) pp.883-890.

Van Den Noort, S. & Holland, N. (1999) *Multiple Sclerosis in Clinical Practice*. New York: Demos Medical Publishing.

Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y. & Fontaine, J. (2008) 'The factor structure of the Beck Depression Inventory-II: an evaluation', *Assessment,* 15 pp.177-187.

Van Overschelde, J. (2008) 'Metacognition: knowing about knowing' in: Dunlosky, J. & Bjork, R. (eds) *Handbook of Metamemory and Memory*. Hove: Psychology Press, pp.47-71.

Vauth, R., Corrigan, P., Clauss, M., Dietl, M., Dreher-Rudolph, M., Stieglitz, R. & Vater, R. (2005) 'Cognitive strategies versus self-management skills as adjunct to vocational rehabilitation', *Schizophrenia Bulletin,* 31 (1) pp.55-66.

Veenman, M., Van Hout-Wolters, B. & Afflerbach, P. (2006) 'Metacognition and learning: conceptual and methodological considerations', *Metacognition and Learning,* 1 (1) pp.3-14.

Veenman, M. & Spaans, M. (2005) 'Relations between intellectual and metacognitive skills: Age and task differences. *Learning & Individual Differences*, 15 pp. 159-176

Veenman, M., Kok, R., & Blote, A. (2005) 'Relation between intellectual and metacognitive skills in early adolescence' *Instructional Science.* 33 (3), 193-211.

Verhaeghen, P., & Salthouse, T. (1997) 'Meta-Analysis of Age-Cognition Relations in Adulthood: Estimates of Linear and Non-Linear Age Effects and Structural Models'. *Psychological Bulletin*, 122 (3) pp. 231-249.

Vikman, T., Fielding, P., Lindmark, B. & Fredrikson, S. (2008) 'Effects of inpatient rehabilitation in multiple sclerosis patients with moderate disability', *Advances in Physiotherapy,* 10 (2) pp.58-65.

Viljoen, J., Iverson, G., Griffiths, S. & Woodward, T. (2003) 'Factor structure of the Beck Depression Inventory-II in a medical outpatient sample', *Journal of Clinical Psychology in Medical Settings,* 10 (4) pp.289-291.

Vilkki, J., Servo, A. & Surma-Aho, O. (1998) 'Word list learning and prediction of recall after frontal lobe lesions', *Neuropsychology*, 12 (2) 268-277.

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

Vriens, M. & Melton, E. (2002) 'Managing missing data multiple imputation can improve data quality', *Marketing Research*, 14 pp.12-17.

Vriens, M. & Sinharay, S. (2006) 'Dealing with missing data in surveys and databases', in: Grover, R. & Vriens, M. (eds) *The Handbook of Marketing Research: Uses, Misuses, and Future Advances.* Thousand Oaks: Sage Publications, pp.178-194.

Wachowius, U., Talley, M., Silver, N., Heinze, H. & Sailer, M. (2005) 'Cognitive impairment in primary and secondary progressive multiple sclerosis', *Journal of Clinical and Experimental Neuropsychology*, 27 (1) pp.65-77.

Wade, D. (1992) 'Stroke: rehabilitation and long-term care', *The Lancet,* 339 (8796) pp.791-793.

Wade, D. (1997) 'Epidemiology of disabling neurological disease: how and why does disability occur?' *Journal of Neurology, Neurosurgery & Psychiatry,* 63 (Suppl 1) pp.S11-S18.

Wade, D. & De Jong, B. (2000) 'Recent advances in rehabilitation', *British Medical Journal,* 320 (7246) pp.1385-1388.

Wang, L., Labar, K., Smoski, M., Rosenthal, M., Dolcos, F., Lynch, T., Krishnan,R. &

Watkins, M. (1990) 'Mediationism and the obfuscation of memory', *The American Psychologist*, 45 (3) pp.328-335.

Watkins, M. (1996) 'Mediationism and the obfuscation of memory', *The Behaviour Analyst,* 19 (1) pp.91-104.

Watkins, E. & Brown, R. (2002) 'Rumination and executive function in depression: an experimental study', *British Medical Journal*, 72 (3) pp.400-402.

Wayman, J. (2003) 'Multiple imputation for missing data: what is it and how can I use it', paper presented at the 2003 *Annual Meeting of the American Educational Research Association*, Chicago.

Wechsler, D. (1997a) *WAIS-III Administration and Scoring Manual.* San Antonio: The Psychological Corporation.

Wechsler, D. (1997b) *Wechsler Adult Intelligence Scale-III (WAIS-III).* New York: Psychological Corporation.

Weinshenker, B., Bass, B., Rice, G., Noseworthy, J., Carriere, W., Baskerville, J. & Ebers, G. (1989) 'The natural history of multiple sclerosis: a geographically based study.I. Clinical course and disability', *Brain*, 112 (1) pp.133-146.

Weiss, D., Shanteau, J. & Harries, P. (2006) 'People who judge people', *Journal of Behavioral Decision Making*, 19 (5) pp.441-454.

Welsh, M., Satterlee-Cartmell, T. & Stine, M. (1999) 'Towers of Hanoi and London: contribution of working memory and inhibition to performance', *Brain and Cognition*, 41 (2) pp.231-242.

West, R. & Berry, J. (1994) 'Age declines in memory self-efficacy: general or limited to particular tasks and measures', in: Sinnott. J. (ed) *Interdisciplinary Handbook of Adult Lifespan Learning.* Westport: Greenwood Press pp.426-445.

West, S., Finch, J. & Curran, P. (1995) 'Structural equation models with nonnormal variables: problems and remedies', in: Hoyle, R. (ed) *Structural Equation Modeling: Concepts, Issues, and Applications.* Thousand Oaks: Sage Publications, pp.56-75.

Weston, R. & Gore Jr, P. (2006) 'A brief guide to structural equation modeling', *The Counseling Psychologist*, 34 (5) pp.719-751.

Wheeler, M. (2000) 'Episodic memory and autonoetic awareness', in: Tulving, E. & Craik, F. (eds) *The Oxford Handbook of Memory*, New York: Oxford University Press, pp.597-608.

Wheeler, M., Stuss, D. & Tulving, E. (1997) 'Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness', *Psychological Bulletin*, 121 (3) pp.331-354.

Whisman, M., Perez, J. & Ramel, W. (2000) 'Factor structure of the Beck Depression Inventory- second edition (BDI-II) in a student sample', *Journal of Clinical Psychology*, 56 (4) pp.545-551.

Wiggins, R. & Sacker, A. (2002) 'Strategies for handling missing data in SEM: a user's perspective', in: Marcoulides, G. & Moustaki, I. (eds) *Latent Variable and Latent Structure Models*, Mahwah: Lawrence Erlbaum, pp.105-121.

Williams, J., O'Rourke, K., Hutchinson, M. & Tubridy, N. (2006) 'The Face-Symbol Test and the Symbol-Digit Test are not reliable surrogates for the Paced Auditory Serial Addition Test in multiple sclerosis', *Multiple Sclerosis*, 12 (5) pp.599-604.

Wilson, B. (1997) 'Cognitive rehabilitation: how it is and how it might be', *Journal of the International Neuropsychological Society*, 3 (5) pp.487-496.

Wilson, B. (2000) 'Compensating for cognitive deficits following brain injury', *Neuropsychology Review*, 10 (4) pp.233-243.

Wilson, B. (2002) 'Towards a comprehensive model of cognitive rehabilitation', *Neuropsychological Rehabilitation*, 12 (2) pp.97-110.

Wilson, B., Clare, L., Cockburn, J., Baddeley, A., Tate, R. & Watson, P. (1999) *The Rivermead Behavioural Memory Test-Extended Version*. Bury St.Edmunds: Thames Valley Test Company.

Wilson, B., Evans, J., Alderman, N., Burgess, P. & Emslie, H. (1997) 'Behavioural assessment of the dysexecutive syndrome', in: Rabbitt, P. (ed) *Methodology of Frontal and Executive Function.* Hove: Psychology Press, pp.239-250.

Wilson, B. & Moffat, N. (eds). (1992) *Clinical Management of Memory Problems*. 2nd edn. London: Chapman & Hall.

Wingerchuk, D. & Weinshenker, B. (2008) 'Neuromyelitis optica', *Current Treatment Options in Neurology,* 10 (1) pp.55-66.

Winstein, C., Wing, A. & Whitall, J. (2003) 'Motor control and learning principles for rehabilitation of upper limb movements after brain injury', in: Grafman, F. & Robertson, I. (eds) *Handbook of Neuropsychology: Plasticity and Rehabilitation Volume 9*. 2nd edn. Amsterdam: Elsevier Science, pp.79-138.

Wright, D. (1996a) 'Measuring feeling of knowing', *Applied Cognitive Psychology,* 10 (3) pp.261-268.

Wright, D. (1996b) 'Measuring feeling of knowing: comment on Schraw (1995)', *Applied Cognitive Psychology,* 10 (3) pp261-268.

Wright, S. (1923) 'The theory of path coefficients: a reply to Niles' criticism', *Genetics,* 8 (3) pp.239-255.

Wykes, T. & Reeder, C. (2005) *Cognitive Remediation Therapy for Schizophrenia: Theory And Practice*, Hove: Routledge.

Ylvisaker., M & Feeney, M., (2004) cited in: Kennedy, M., (2006) Managing Memory and Metammeory Impairments in Individuals with Traumatic brain Injury. The ASHA leader. *American Speech-Language-Hearing Association* online: accessed February 22, 2010 [http://www.asha.org/Publications/leader/2006/061017/f061017b.htm]

Young, R., (2000) 'Spastic Paresis', in: Burks, J. & Johnson, K. (eds) *Multiple Sclerosis: Diagnosis, Medical Management, and Rehabilitation*. New York: Demos Medical Publishing, pp.299-306.

Yuan, K. (2005) 'Fit indices versus test statistics', *Multivariate Behavioral Research,* 40 (1) pp.115-148.

Yuan, K. & Hayashi, K. (2003) 'Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models' *British Journal of Mathematical and Statistical Psychology*, 56 (1) pp. 93-110.

Zelinski, E. & Gilewski, M. (2004) 'A 10-item Rasch modeled memory self-efficacy scale', *Aging and Mental Health,* 8 (4) pp.293-306.

Zelinski, E., Gilewski, M. & Anthony-Bergstone, C. (1990) 'Memory Functioning Questionnaire: concurrent validity with memory performance and self-reported memory failures', *Psychology and Aging*, 5 (3) pp.388-399.

Zook, N., Davalos, D., Delosh, E. & Davis, H. (2004) 'Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks', *Brain and Cognition*, 56 (3) pp.286-292.

This list is indicative of the materials used, and not the layout of the stimulus items. The practice item is first, followed by the 24 sentences and original completion word. Listed below each sentence are the 8 words presented for the 7-alternative forced choice recognition test.

She put the flowers in the VASE

GARDEN
WATER
KITCHEN
COFFIN
SOIL
VASE
BIN
WINDOW

There is something grand about the HOUSE

HOTEL
OPERA
BUILDING
HOUSE
PIANO
PALACE
WEATHER
FLAG

Their money was divided by the TEACHER

PRIEST
TEACHER
SOLICITOR
BANK
FAMILY
JUDGE
COURT
FATHER

There was nothing wrong with the MEAL

COMPUTER
RADIO
MEAL
CAR
TELEVISION
FOOD
MACHINE
DINNER

I don’t know why he didn’t take his UMBRELLA

MEDICINE
UMBRELLA
TABLETS
KEYS
HAT
COAT
MONEY
GLOVES

They went to the rear of the long CORRIDOR

BOAT
QUEUE
ROOM
CARRIAGE
TRAIN
BUS
HALL
CORRIDOR

He wondered if the storm would be VIOLENT

FIERCE
VIOLENT
OVER
LONG
HEAVY
BADMILD
DANGEROUS

They went to see the famous PAINTING

FIVE
PAINTING
ACTOR
STATUE
BUILDING
CHURCH
FILM
SINGER

Did you want to go to the CITY

PARK
CITY
CINEMA
PICTURES
TOILET
SHOPS
PUB
BALL

The judge warned about the dangers of LYING

SPEED
LYING
DRUGS
DRINKING
CRIME
ROBBERY
ALCOHOL
REVENGE

She dropped a glass and woke up her CAT

MOTHER
CAT
BABY
HUSBAND
SISTER
FRIEND
DOG
NEIGHBOUR

Rita slowly walked down the shaky PLANK

DECK
PLANK
STAIRS
LADDER
BRIDGE
STAIRCASE
PATH
TREE

Larry chose not to join the GANG

CLUB
GANG
ARMY
SCOUTS
GAME
NAVY
FIGHT
TEAM

Hank reached into his pocket to get the LIGHTER

TICKET
LIGHTER
MONEY
KEYS
GUN
CHANGE
PEN
CIGARETTES

Ray fell down and hurt his BACK

PRIDE
BACK
KNEE
ARM
ELBOW
HAND
ANKLE
HIP

Suzy liked to play with her toy BEAR

DOG
PHONE
DOLLS
SOLDIERS
DESK
TRAIN
BEAR
DRUM

The sandwich wasn’t very good without a slice of TOMATO

TOMATO
BREAD
CHEESE
HAM
MEAT
ONION
CUCUMBER
PICKLE

She cleaned the dirt from her FACE

SHOES
FACE
NAILS
BOOTS
HANDS
COAT
HAIR
GLASSES

Rushing out, he forgot to take his WALLET

MONEY
WALLET
KEYS
COAT
BAG
JACKET
TICKETS
BOOKS

Helen reached up to dust the MANTELPIECE

MANTELPIECE
SHELVES
WARDROBE
PICTURE
MIRROR
CEILING
CUPBOARD
LAMP

James poured himself a glass of WHISKY

ORANGE
WHISKY
WINE
BEER
MILK
JUICE
SHERRY
LEMONADE

Her dress was made of very fine LACE

MATERIAL
LACE
SILK
CLOTH
LINEN
COTTON
FABRIC
SEWING

The lorry that Bill drove crashed into the BRIDGE

BARRIER
BRIDGE
WALL
HOUSE
CAR
TRAIN
LIGHTS
ROAD

The hunter shot and killed a large ELEPHANT

ELEPHANT
DEER
BOAR
LION
TIGER
RABBIT
BIRD
ANIMAL

The police had never seen a man so ANGRY

NEGLECTED
ANGRY
DRUNK
VIOLENT
POOR
ANXIOUS
GUILTY
INNOCENT

Brunel
UNIVERSITY
WEST LONDON

School of Health Sciences and Social Care

Research Ethics Committee

School of Health Sciences and
Social Care
Brunel University,
Uxbridge
Middlesex UB8 3PH
Telephone: +44 (0)1895 274000
Web www.brunel.ac.uk

19 September 2007

Proposer: **Austin Claffey – PhD Student**

Title: **An Investigation of Factors Contributing to Metamemory in Multiple Sclerosis**

Reference: **07/07/PHD/06**

Letter of Approval

The School Research Ethics Committee has considered the amendments recently submitted by you in response to the Committee's earlier review of the above application

The Chair, acting under delegated authority, is satisfied that the amendments accord with the decision of the Committee and has agreed that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that the conditions of approval set out below are followed:

- *The agreed protocol must be followed. Any changes to the protocol will require prior approval from the Committee.*

NB:

- **Research participant information sheets and (where relevant) flyers, posters and consent forms, should include a clear statement that research ethics approval has been obtained from the School of Health Sciences and Social Care Research Ethics Committee.**
- **Approval to proceed with the study is granted subject to receipt by the Committee of satisfactory responses to any conditions that may appear above, in addition to any subsequent changes to the protocol.**

David Anderson-Ford
Chair, Research Ethics Committee
School of Health Sciences and Social Care

Brunel UNIVERSITY WEST LONDON

Research Participant Information Sheet.

Research Title: Knowledge of Memory Ability in Multiple Sclerosis

Researcher's name and contact details:

Mr. Austin Claffey, PhD student, Brunel University
Telephone number: 07504517954
Email: austin.claffey@brunel.ac.uk

Research Supervisors:

Dr Priscilla Harries & Prof. Lorraine DeSouza,
School of Health Sciences & Social Care,
Mary Seacole Building,
Brunel University,
Uxbridge,
Middlesex, UB8 3PH.

This research is a PhD study being undertaken at Brunel University, London. This study has been granted ethical approval by the Brunel University Research Ethics Committee.

The information sheet gives an explanation of the research and what would happen to you if you agreed to be involved. If you have any questions about the project, please make contact with the researcher.

What is the research about?

This research project aims to understand how people with multiple sclerosis estimate their memory ability. People with Multiple Sclerosis sometimes report memory problems. Some researchers have suggested that problems such as concentration, information processing or your mood can play a part. This study will investigate what are the issues associated with reports of memory problems. **You do not have to be having memory problems to take part in this study.**

What will the study involve?

I would like you to have one meeting with me, consisting of two sessions separated by a break of about 20 minutes. Each session will be 30-40 minutes long. The sessions will consist of taking some background information about you and your medical history, followed by testing of some aspects of your memory, thinking speed and reasoning. There will be a short questionnaire about your mood at the end. In order to test speed some of your tests answers will be recorded on tape so that they can be timed.

Before attending for the assessment session, you will be asked to fill out one questionnaire about your memory, which I will post to you in advance, so you can bring it along with you.

There will not be any treatments given to you during the research, nor will there be a need to stop taking medication. Times and dates for assessment will be organised to fit in whenever suits you best, and the meeting can be at your MS Therapy Centre, Brunel University or in your own home, whichever you prefer.

What are the criteria for getting involved?

If you are over 18 years old, living in the community, and have a doctor's diagnosis of Multiple Sclerosis, I would be interested in hearing from you. The research tasks will require some basic reading, writing, pointing and speaking. Finally, you will need to have no neurological problems apart from MS, and no history of depression prior to your diagnosis of MS. If you have had a recent exacerbation of your MS, this will need to be discussed, but I would still like you to take part.

What are the benefits of getting involved?

The results of this study may help people with MS and clinicians better understand what contributes to the ability to estimate memory ability. It is unlikely that there will be a direct benefit to you from this, during the study. It is however hoped that if we have a better understanding of what contributes to the experience of memory problems, we will better be able to support rehabilitation. If during the study it is identified that you are having significant problems with your mood, I may recommend to you that you see your GP for further advice or treatment.

Results from this study will be submitted for publication, or presented at conferences, so that they can reach an audience of professionals, as well as people with MS.

What are the risks of getting involved?

One section of the study will ask you to fill in a questionnaire about your mood over the past two weeks. This questionnaire sometimes provokes feelings of upset because it focuses on a range of negative thoughts and feelings that some people experience. Other testing components are designed to be sensitive to speed of thinking or other mental abilities. These tests are designed so that they challenge these abilities and, for some, this can cause some anxiety. This anxiety is considered normal in a testing situation.

Do I have to get involved?

No, your participation is voluntary. This means that you can choose not to be involved, or if you start, you can choose to withdraw at any time. If you decide not to participate, it will not affect any other treatments you are receiving.

Will my details and results be confidential?

Yes, all the information you provide will be kept confidential. Any *personal information* will be kept securely in the university and anonymous information, like assessment scores, will be kept separate from it.

Any information that could identify you will be removed before the research is analysed, written up, or before any presentation of the results is made.

Who is supervising the research?

The research plan has been reviewed and accepted by Brunel University Ethics Committee and the university also funds and insures the researcher. If you have are unhappy with your experience or have complaints about the conduct of the research you can contact the supervisor of the research.

<u>The contact details of the project supervisor are:</u>

Dr Priscilla Harries,

School of Health Sciences & Social Care,

Mary Seacole Building,

Brunel University,

Uxbridge,

Middlesex, UB8 3PH. Telephone: 01895 268773

Will I be paid to take part?

Any travel expenses you incur as a result of participating in the research will be refunded. An honorarium of a £15 gift token will be given to you at the end of the testing session, to thank you for giving me your time.

Will you contact anyone else about my participation?

If you would like me to, I can discuss the study with a family member, partner, friend or carer. Professional staff, such as therapists, nurses or care workers will not be informed about your results, unless you ask for them to be.

What happens next?

If you decide to take part, I will ask you to read and sign a consent form. A copy is enclosed with this information sheet so you can have a read of it. If you have any questions about the study at this point I can answer them for you.

After that, two things will happen:

1. You will be given one questionnaire to fill out, which asks you about your memory. This will take about 30 minutes to fill in.
2. An appointment will be arranged for you for the assessment part of the research, which will consist of some memory, thinking speed and reasoning assessments. This will take about 90 minutes to do, including a break. A questionnaire about your mood will be included in this session.

Where will the appointment be?

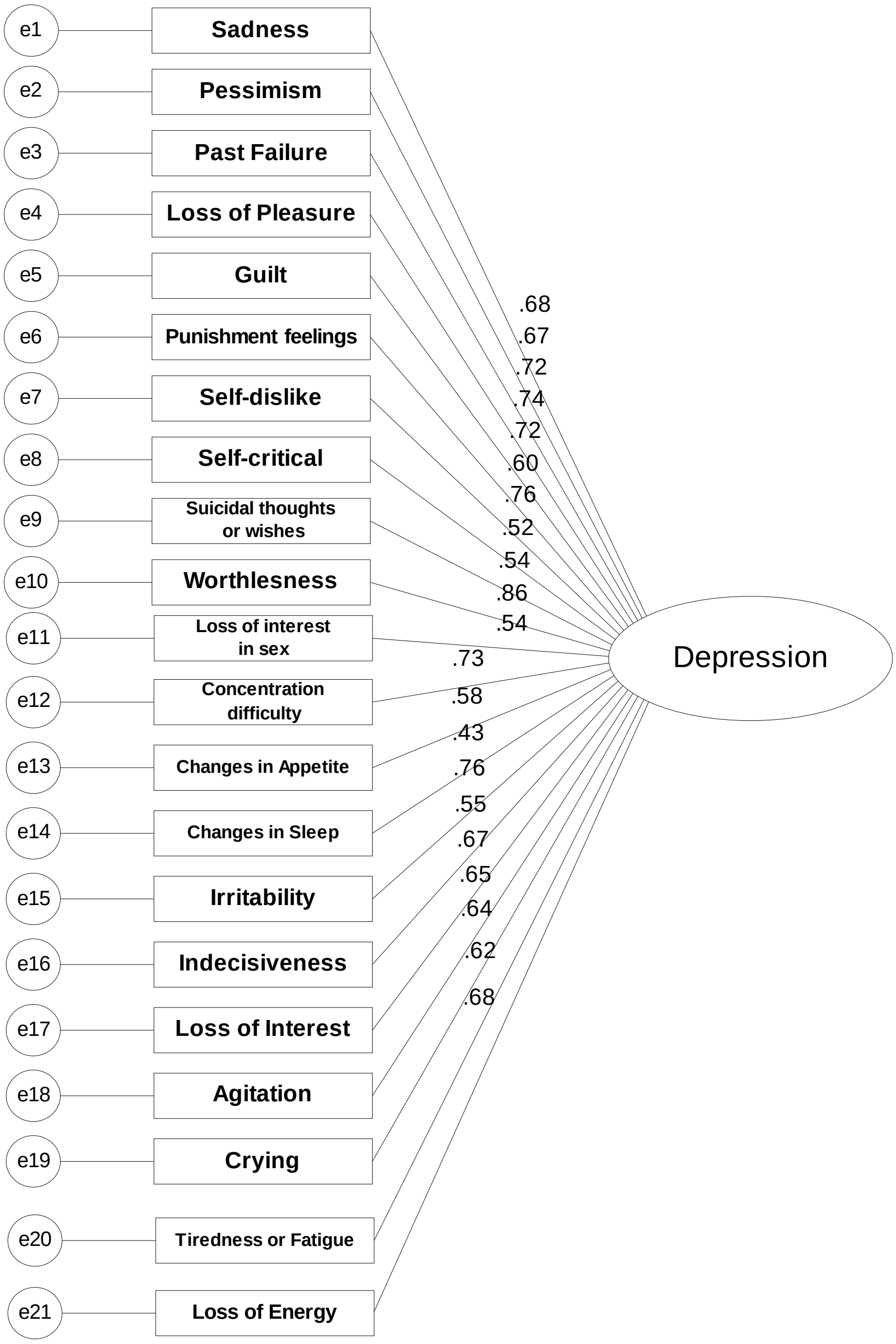
It could be at your home, at the university or possibly at the centre you attend. The researcher will discuss this with you.

Thank you for considering being part of this study.

If you have any other questions about the project, please make contact with the researcher:

Austin Claffey
Telephone number: 07504517954
Email: austin.claffey@brunel.ac.uk

Appendix D: 1 FACTOR CFA of BECK DEPRESSION INVENTORY



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables
Loss of Pleasure
Guilt Feelings
Punishment Feelings
Worthlessness
Suicidal Thoughts or wishes
Self-Critical
Self-Dislike
Sadness
Pessimism
Past Failure
Loss of interest in Sex
Concentration difficulty
Change in Appetite
Change in Sleep
Irritated
Indecisiveness
Loss of Interest
Agitation
Crying
Tiredness or Fatigue
Loss of Energy

Unobserved, exogenous variables
Depression

Variable counts (including residuals)

Number of variables in your model:	43
Number of observed variables:	21
Number of unobserved variables:	22
Number of exogenous variables:	22
Number of endogenous variables:	21

Computation of degrees of freedom

Number of distinct sample moments:	231
Number of distinct parameters to be estimated:	42
Degrees of freedom (231 - 42):	189

Result

Minimum was achieved
Chi-square = 233.653
Degrees of freedom = 189
Probability level = .015

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
BDI1factor model	42	233.653	189	.015	1.236
Saturated model	231	.000	0		
Independence model	21	278.161	210	.001	1.325
Zero model	0	1039.50	231	.000	4.500

GFI

Model	GFI
BDI1factor model	.775
Saturated model	1.00
Independence model	.732
Zero model	.000

Baseline Comparisons

Model	CFI
BDI1factor model	.345
Saturated model	1.00
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
BDI1factor model	.049	.023	.068	.523
Independence model	.057	.037	.075	.253

Regression Weights:

			Estimate	S.E.	C.R.	P
Loss of Pleasure	<---	Depression	1.493	.512	2.915	.004
Guilt Feelings	<---	Depression	1.277	.307	4.160	***
Punishment Feelings	<---	Depression	1.272	.447	2.847	.004
Worthlessness	<---	Depression	1.810	.551	3.283	.001
Suicidal Thoughts or wishes	<---	Depression	.806	.327	2.468	.014
Self-Critical	<---	Depression	1.054	.349	3.018	.003
Self-Dislike	<---	Depression	1.383	.421	3.283	.001
Sadness	<---	Depression	1.000			
Pessimism	<---	Depression	1.319	.446	2.955	.003
Past Failure	<---	Depression	1.339	.471	2.841	.005
Loss of interest in Sex	<---	Depression	1.138	.526	2.166	.030
Concentration difficulty	<---	Depression	1.435	.493	2.911	.004
Change in Appetite	<---	Depression	1.172	.462	2.538	.011
Change in Sleep	<---	Depression	.903	.434	2.080	.038
Irritated	<---	Depression	1.221	.388	3.149	.002
Indecisiveness	<---	Depression	1.157	.469	2.465	.014
Loss of Interest	<---	Depression	1.079	.434	2.487	.013
Agitation	<---	Depression	1.152	.359	3.203	.001
Crying	<---	Depression	1.149	.477	2.408	.016
Tiredness or Fatigue	<---	Depression	1.168	.475	2.458	.014
Loss of Energy	<---	Depression	1.077	.457	2.359	.018

Standardized Regression Weights:

			Estimate
Loss of Pleasure	<---	Depression	.745
Guilt Feelings	<---	Depression	.724
Punishment Feelings	<---	Depression	.595
Worthlessness	<---	Depression	.857
Suicidal Thoughts or wishes	<---	Depression	.541
Self-Critical	<---	Depression	.517
Self-Dislike	<---	Depression	.764
Sadness	<---	Depression	.676
Pessimism	<---	Depression	.672
Past Failure	<---	Depression	.721
Loss of interest in Sex	<---	Depression	.537
Concentration difficulty	<---	Depression	.728
Change in Appetite	<---	Depression	.578
Change in Sleep	<---	Depression	.429
Irritated	<---	Depression	.759
Indecisiveness	<---	Depression	.552
Loss of Interest	<---	Depression	.669
Agitation	<---	Depression	.651
Crying	<---	Depression	.638

			Estimate
Tiredness or Fatigue	<---	Depression	.620
Loss of Energy	<---	Depression	.680

Variances

	Estimate	S.E.	C.R.	P
Depression	.061	.031	1.954	.051
e1	.073	.015	4.830	***
e2	.130	.025	5.193	***
e3	.101	.023	4.459	***
e4	.109	.023	4.682	***
e5	.091	.020	4.576	***
e6	.180	.032	5.557	***
e7	.084	.020	4.252	***
e8	.186	.031	6.051	***
e9	.096	.018	5.490	***
e10	.072	.016	4.628	***
e11	.196	.035	5.630	***
e13	.168	.031	5.453	***
e12	.112	.021	5.415	***
e14	.221	.037	5.891	***
e15	.067	.015	4.535	***
e16	.187	.035	5.351	***
e17	.088	.020	4.330	***
e18	.110	.021	5.264	***
e19	.118	.025	4.786	***
e20	.134	.027	5.021	***
e21	.083	.017	4.900	***

Squared Multiple Correlations:

	Estimate
Agitation	.424
Change in Appetite	.334
Change in Sleep	.530
Concentration difficulty	.268
Crying	.407
Guilt Feelings	.583
Indecisiveness	.463
Irritated	.520
Loss of Energy	.384
Loss of Interest	.524
Loss of interest in Sex	.305
Loss of Pleasure	.448
Past Failure	.577
Pessimism	.451
Punishment Feelings	.555
Sadness	.354
Self-Critical	.456
Self-Dislike	.289
Suicidal Thoughts or wishes	.184
Tiredness or Fatigue	.292
Worthlessness	.735

Sample Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Energy	.223																				
Fatigue	.113	.340																			
Cry	.071	.071	.312																		
Agitate	.064	.100	.111	.268																	
Interest	.064	.076	.096	.086	.263																
Indecisive	.109	.079	.088	.083	.075	.380															
Irritate	.061	.073	.092	.114	.117	.067	.214														
Sleep	.088	.076	.122	.077	.053	.098	.070	.365													
Appetite	.093	.121	.128	.112	.039	.077	.110	.081	.353												
Concentrate	.132	.163	.128	.111	.095	.097	.104	.105	.133	.303											
Sex	.070	.143	.101	.086	.080	.106	.091	.124	.132	.111	.390										
Fail	.067	.039	.058	.060	.031	.085	.088	.071	.080	.094	.076	.329									
Pess	.107	.060	.074	.045	.094	.065	.086	.027	.061	.126	.039	.108	.329								
Sad	.030	.072	.061	.094	.065	.065	.067	.044	.066	.079	.027	.089	.091	.195							
Dislike	.079	.099	.089	.093	.052	.148	.098	.080	.080	.112	.085	.075	.097	.086	.282						
Critical	.067	.059	.089	.073	.104	.066	.083	.036	.041	.111	.031	.058	.121	.070	.085	.316					
Suicide	.061	.057	.033	.003	.053	.025	.063	.025	.013	.046	.017	.081	.102	.049	.059	.060	.196				
Worth	.120	.083	.128	.102	.104	.135	.115	.068	.102	.135	.108	.203	.156	.091	.147	.102	.104	.321			
Punish	.059	.123	.090	.048	.099	.063	.054	.054	.095	.093	.141	.092	.138	.075	.103	.078	.073	.147	.376		
Guilt	.053	.091	.044	.093	.048	.052	.087	.042	.090	.069	.086	.136	.069	.123	.129	.101	.065	.134	.109	.284	
Pleas	.088	.062	.031	.098	.107	.075	.104	.061	.076	.091	.088	.146	.142	.073	.138	.073	.101	.186	.124	.118	.340

Sample Correlations

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Energy	1.000																				
Fatigue	.411	1.000																			
Cry	.270	.217	1.000																		
Agitate	.263	.331	.385	1.000																	
Interest	.265	.253	.334	.323	1.000																
Indecisive	.375	.219	.254	.259	.238	1.000															
Irritate	.277	.270	.355	.475	.492	.233	1.000														
Sleep	.307	.215	.361	.245	.172	.264	.251	1.000													
Appetite	.332	.350	.385	.364	.127	.211	.398	.226	1.000												
Concentrate	.506	.507	.416	.389	.336	.287	.407	.315	.406	1.000											
Sex	.236	.392	.289	.265	.250	.275	.315	.328	.355	.322	1.000										
Fail	.246	.117	.180	.203	.104	.240	.333	.204	.234	.298	.212	1.000									
Pess	.396	.179	.232	.153	.320	.185	.322	.077	.180	.398	.109	.328	1.000								
Sad	.143	.280	.249	.409	.286	.237	.329	.165	.251	.323	.098	.350	.360	1.000							
Dislike	.313	.320	.299	.337	.191	.452	.401	.250	.255	.383	.257	.247	.319	.368	1.000						
Critical	.253	.180	.283	.251	.360	.192	.318	.106	.122	.357	.089	.180	.374	.282	.285	1.000					
Suicide	.290	.222	.135	.013	.235	.093	.306	.092	.051	.188	.061	.318	.402	.251	.252	.242	1.000				
Worth	.447	.253	.406	.346	.356	.386	.439	.197	.304	.432	.306	.624	.480	.365	.490	.321	.415	1.000			
Punish	.202	.343	.262	.152	.314	.168	.190	.145	.260	.274	.369	.262	.391	.275	.315	.226	.268	.424	1.000		
Guilt	.210	.293	.146	.338	.175	.158	.354	.130	.283	.236	.257	.444	.226	.520	.454	.338	.275	.444	.332	1.000	
Pleas	.321	.182	.094	.326	.359	.208	.386	.174	.220	.285	.243	.437	.426	.285	.446	.222	.390	.565	.346	.381	1.000

Matrices

Residual Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Energy	.070																				
Fatigue	.036	.122																			
Cry	-.004	-.011	.113																		
Agitate	-.012	.017	.030	.077																	
Interest	-.007	-.002	.020	.010	.104																
Indecisive	.033	-.004	.006	.001	-.001	.112															
Irritate	-.020	-.014	.006	.028	.036	-.020	.056														
Sleep	.028	.011	.058	.013	-.006	.034	.003	.094													
Appetite	.016	.037	.045	.029	-.039	-.006	.022	.016	.102												
Concentrate	.037	.060	.027	.010	.000	-.004	-.004	.026	.030	.065											
Sex	-.005	.061	.021	.006	.005	.025	.006	.061	.050	.011	.115										
Fail	-.022	-.057	-.037	-.034	-.058	-.010	-.012	-.003	-.016	-.024	-.017	.118									
Pess	.020	-.035	-.018	-.048	.007	-.028	-.013	-.046	-.033	.010	-.053	.000	.093								
Sad	-.036	.001	-.009	.023	-.001	-.006	-.008	-.011	-.006	-.009	-.043	.007	.010	.061							
Dislike	-.013	.000	-.009	-.005	-.039	.050	-.005	.004	-.019	-.010	-.011	-.038	-.015	.002	.081						
Critical	-.002	-.017	.015	-.001	.034	-.008	.004	-.022	-.035	.018	-.042	-.028	.035	.005	-.004	.062					
Suicide	.007	.000	-.023	-.054	.000	-.032	.002	-.020	-.044	-.025	-.039	.015	.037	.000	-.009	.008	.060				

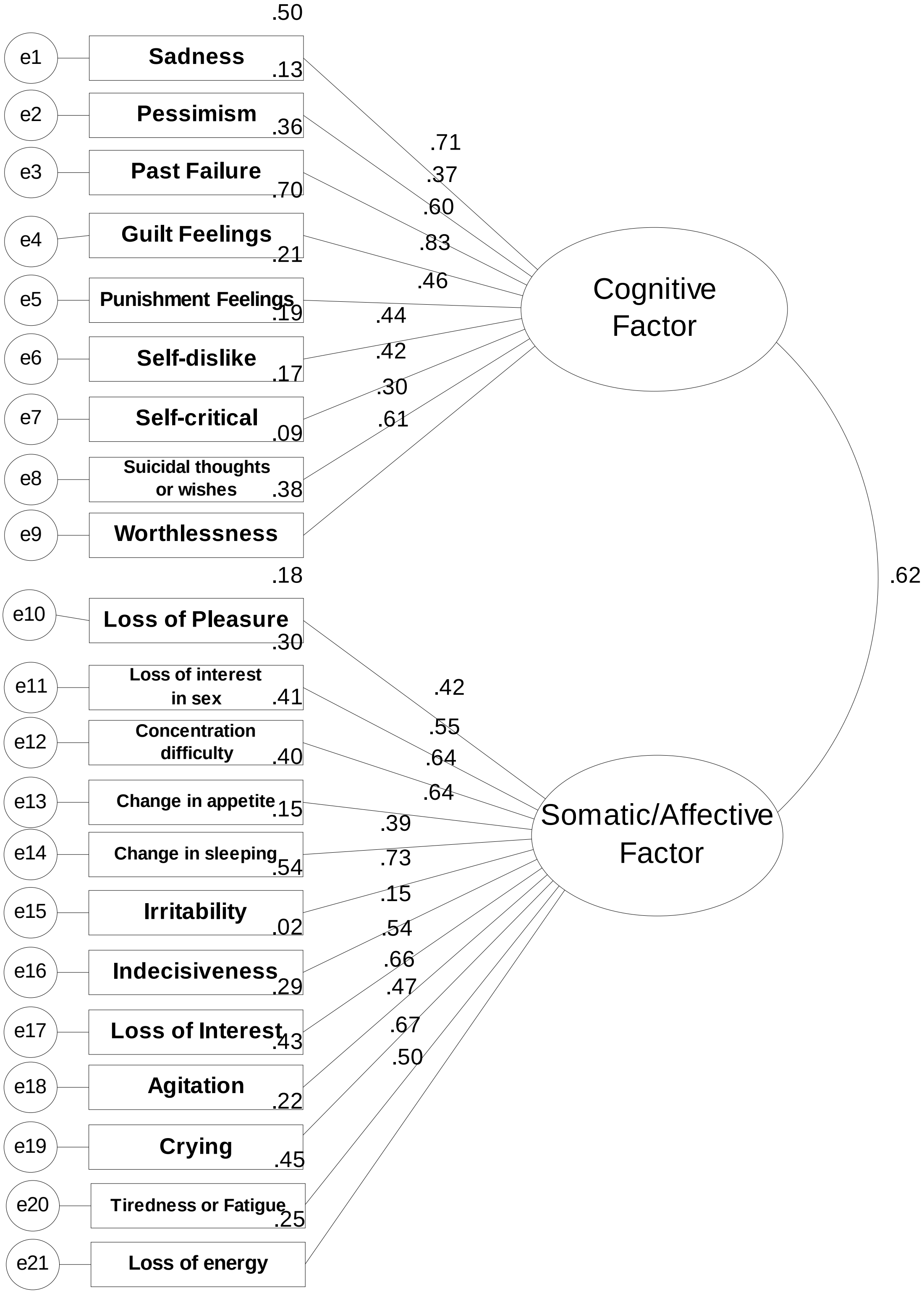
	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Worth	.000	-.046	.001	-.026	-.016	.007	-.020	-.033	-.028	-.024	-.018	.054	.010	-.019	-.006	-.015	.015	.048			
Punish	-.025	.032	.000	-.041	.015	-.027	-.041	-.017	.004	-.019	.053	-.012	.035	-.003	-.005	-.004	.010	.007	.097		
Guilt	-.031	.000	-.046	.003	-.036	-.039	-.008	-.029	-.002	-.043	-.003	.031	-.034	.044	.021	.019	.002	-.007	.009	.094	
Pleas	-.010	-.045	-.074	-.007	.009	-.031	-.007	-.021	-.031	-.040	-.016	.024	.022	-.018	.012	-.024	.027	.021	.007	.002	.094

Standardized Residual Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Energy	3.193																				
Fatigue	1.818	3.941																			
Cry	-.235	-.509	4.014																		
Agitate	-.614	.789	1.428	2.810																	
Interest	-.405	-.075	1.027	.511	4.603																
Indecisive	1.514	-.153	.256	.054	-.053	2.923															
Irritate	-1.129	-.703	.302	1.416	2.014	-.891	2.478														
Sleep	1.313	.443	2.416	.545	-.299	1.234	.118	2.441													
Appetite	.746	1.495	1.893	1.247	-1.792	-.207	.999	.610	2.837												
Concentrate	1.728	2.398	1.118	.415	.012	-.154	-.165	.957	1.122	1.931											
Sex	-.245	2.360	.833	.227	.221	.883	.257	2.163	1.810	.387	2.949										
Fail	-1.079	-2.406	-1.616	-1.522	-2.823	-.379	-.567	-.137	-.646	-.929	-.669	3.923									
Pess	.958	-1.399	-.775	-2.044	.325	-1.040	-.603	-1.752	-1.271	.363	-1.949	-.016	2.757								

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc.	Sex	Fail	Pess	Sad	Dislike	Critical	Suicide	Worth	Punish	Guilt	Pleas
Sad	-2.277	.032	-.497	1.318	-.078	-.309	-.460	-.563	-.290	-.466	-2.084	.356	.529	3.210							
Dislike	-.638	.008	-.382	-.215	-1.948	1.973	-.241	.151	-.762	-.383	-.443	-1.619	-.600	.088	2.840						
Critical	-.117	-.666	.620	-.049	1.593	-.301	.183	-.821	-1.313	.680	-1.536	-1.136	1.361	.277	-.171	1.703					
Suicide	.478	-.027	-1.331	-3.130	.000	-1.585	.155	-1.013	-2.277	-1.287	-1.940	.799	1.927	-.012	-.508	.414	3.083				
Worth	.014	-1.661	.043	-.988	-.663	.218	-.809	-1.118	-.937	-.805	-.589	1.918	.327	-.871	-.213	-.502	.681	1.236			
Punish	-1.126	1.193	.009	-1.655	.650	-.920	-1.774	-.583	.126	-.679	1.797	-.454	1.256	-.155	-.197	-.152	.484	.209	2.437		
Guilt	-1.628	-.007	-2.147	.152	-1.869	-1.576	-.410	-1.208	-.075	-1.768	-.136	1.366	-1.432	2.476	.914	.793	.114	-.269	.362	3.460	
Pleas	-.460	-1.754	-3.018	-.288	.387	-1.111	-.328	-.777	-1.132	-1.436	-.561	.906	.799	-.885	.448	-.882	1.357	.679	.256	.069	2.684

Appendix E - 2 FACTOR BDI COGNITIVE FACTOR, AFFECTIVE/SOMATIC FACTOR



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables
Loss of Pleasure
Guilt Feelings
Punishment Feelings
Worthlessness
Suicidal Thoughts or wishes
Self-Critical
Self-Dislike
Sadness
Pessimism
Past Failure
Loss of interest in Sex
Concentration difficulty
Change in Appetite
Change in Sleep
Irritated
Indecisiveness
Loss of Interest
Agitation
Crying
Tiredness or Fatigue
Loss of Energy

Unobserved, exogenous variables
Cognitive
Somatic_Affective

Variable counts (including residuals)

Number of variables in your model:	44
Number of observed variables:	21
Number of unobserved variables:	23
Number of exogenous variables:	23
Number of endogenous variables:	21

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	23	0	0	0	0	23
Labeled	0	0	0	0	0	0
Unlabeled	19	1	23	0	0	43
Total	42	1	23	0	0	66

Computation of degrees of freedom

Number of distinct sample moments:	231
Number of distinct parameters to be estimated:	43
Degrees of freedom (231 - 43):	188

Result

Chi-square = 229.131
Degrees of freedom = 188
Probability level = .022

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
BDI2factor	43	229.131	188	.022	1.219
Saturated model	231	.000	0		
Independence model	21	278.161	210	.001	1.325
Zero model	0	1039.500	231	.000	4.500

RMR, GFI

Model	GFI
BDI2factor	.780
Saturated model	1.000
Independence model	.732
Zero model	.000

Baseline Comparisons

Model	CFI
BDI2factor	.397
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
BDI2factor	.047	.019	.067	.578
Independence model	.057	.037	.075	.253

Regression Weights

			Estimate	S.E.	C.R.	P
Guilt	<---	Cognitive	1.522	.312	4.874	***
Punish	<---	Cognitive	.826	.263	3.136	.002
Sleep	<---	Somatic_Affective	.711	.247	2.884	.004
Irritate	<---	Somatic_Affective	.968	.226	4.282	***
Indecisive	<---	Somatic_Affective	.241	.230	1.047	.295
Interest	<---	Somatic_Affective	.691	.212	3.259	.001
Sex	<---	Somatic_Affective	1.035	.276	3.746	***
Concentrate	<---	Somatic_Affective	1.000			
Appetite	<---	Somatic_Affective	1.199	.278	4.312	***
Cry	<---	Somatic_Affective	.659	.213	3.100	.002
Agitate	<---	Somatic_Affective	1.030	.242	4.248	***
Worth	<---	Cognitive	.773	.220	3.514	***
Suicide	<---	Cognitive	.368	.175	2.098	.036
Critical	<---	Cognitive	.748	.246	3.046	.002
Dislike	<---	Cognitive	.525	.203	2.588	.010
Sad	<---	Cognitive	1.000			
Pess	<---	Cognitive	.540	.217	2.490	.013
Fail	<---	Cognitive	.919	.245	3.743	***
Fatigue	<---	Somatic_Affective	1.162	.245	4.747	***
Energy	<---	Somatic_Affective	.600	.168	3.577	***
Pleas	<---	Somatic_Affective	.527	.224	2.352	.019

Covariances: (Group number 1 - BDI2factor)

			Estimate	S.E.	C.R.	P
Cognitive	<-->	Somatic_Affective	.046	.017	2.747	.006

Variances: (Group number 1 - BDI2factor)

	Estimate	S.E.	C.R.	P
Cognitive	.070	.024	2.931	.003
Somatic_Affective	.077	.029	2.638	.008
e1	.070	.016	4.250	***
e2	.131	.025	5.281	***
e3	.103	.023	4.505	***
e5	.071	.026	2.734	.006
e6	.179	.033	5.461	***
e7	.082	.020	4.132	***
e8	.185	.031	5.953	***
e9	.098	.018	5.563	***
e10	.069	.015	4.471	***
e18	.111	.021	5.338	***
e17	.163	.032	5.138	***
e16	.220	.038	5.851	***
e15	.063	.016	4.018	***
e14	.185	.035	5.256	***
e13	.090	.020	4.394	***
e12	.109	.022	5.043	***
e11	.119	.025	4.837	***
e21	.083	.017	4.933	***
e20	.127	.028	4.580	***
e19	.195	.035	5.518	***
e4	.098	.024	4.105	***

Squared Multiple Correlations: (Group number 1 - BDI2factor)

	Estimate
Pleasure	.179
Sex	.298
Fatigue	.450
Energy	.250
Cry	.220
Agitate	.429
Interest	.291
Indecisive	.024
Irritate	.536
Sleep	.151
Appetite	.404
Concentrate	.409
Worth	.375
Suicide	.088
Critical	.174
Dislike	.190
Punish	.210
Guilt	.696
Fail	.363
Pessimism	.134
Sad	.500

Sample Moments

Sample Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	.223																				
Fatigue	.113	.340																			
Cry	.071	.071	.312																		
Agitate	.064	.100	.111	.268																	
Interest	.064	.076	.096	.086	.263																
Indecisive	.109	.079	.088	.083	.075	.380															
Irritate	.061	.073	.092	.114	.117	.067	.214														
Sleep	.088	.076	.122	.077	.053	.098	.070	.365													
Concentrate	.132	.163	.128	.111	.095	.097	.104	.105	.303												
Appetite	.093	.121	.128	.112	.039	.077	.110	.081	.133	.353											
Sex	.070	.143	.101	.086	.080	.106	.091	.124	.111	.132	.390										
Worth	.120	.083	.128	.102	.104	.135	.115	.068	.135	.102	.108	.321									
Suicide	.061	.057	.033	.003	.053	.025	.063	.025	.046	.013	.017	.104	.196								
Critical	.067	.059	.089	.073	.104	.066	.083	.036	.111	.041	.031	.102	.060	.316							
Dislike	.079	.099	.089	.093	.052	.148	.098	.080	.112	.080	.085	.147	.059	.085	.282						
Punish	.059	.123	.090	.048	.099	.063	.054	.054	.093	.095	.141	.147	.073	.078	.103	.376					
Guilt	.053	.091	.044	.093	.048	.052	.087	.042	.069	.090	.086	.134	.065	.101	.129	.109	.284				
Pleas	.088	.062	.031	.098	.107	.075	.104	.061	.091	.076	.088	.186	.101	.073	.138	.124	.118	.340			
Fail	.067	.039	.058	.060	.031	.085	.088	.071	.094	.080	.076	.203	.081	.058	.075	.092	.136	.146	.329		
Pess	.107	.060	.074	.045	.094	.065	.086	.027	.126	.061	.039	.156	.102	.121	.097	.138	.069	.142	.108	.329	
Sad	.030	.072	.061	.094	.065	.065	.067	.044	.079	.066	.027	.091	.049	.070	.086	.075	.123	.073	.089	.091	.195

Sample Correlations

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	1.000																				
Fatigue	.411	1.000																			
Cry	.270	.217	1.000																		
Agitate	.263	.331	.385	1.000																	
Interest	.265	.253	.334	.323	1.000																
Indecisive	.375	.219	.254	.259	.238	1.000															
Irritate	.277	.270	.355	.475	.492	.233	1.000														
Sleep	.307	.215	.361	.245	.172	.264	.251	1.000													
Concentrate	.506	.507	.416	.389	.336	.287	.407	.315	1.000												
Appetite	.332	.350	.385	.364	.127	.211	.398	.226	.406	1.000											
Sex	.236	.392	.289	.265	.250	.275	.315	.328	.322	.355	1.000										
Worth	.447	.253	.406	.346	.356	.386	.439	.197	.432	.304	.306	1.000									
Suicide	.290	.222	.135	.013	.235	.093	.306	.092	.188	.051	.061	.415	1.000								
Critical	.253	.180	.283	.251	.360	.192	.318	.106	.357	.122	.089	.321	.242	1.000							
Dislike	.313	.320	.299	.337	.191	.452	.401	.250	.383	.255	.257	.490	.252	.285	1.000						
Punish	.202	.343	.262	.152	.314	.168	.190	.145	.274	.260	.369	.424	.268	.226	.315	1.000					
Guilt	.210	.293	.146	.338	.175	.158	.354	.130	.236	.283	.257	.444	.275	.338	.454	.332	1.000				
Pleas	.321	.182	.094	.326	.359	.208	.386	.174	.285	.220	.243	.565	.390	.222	.446	.346	.381	1.000			
Fail	.246	.117	.180	.203	.104	.240	.333	.204	.298	.234	.212	.624	.318	.180	.247	.262	.444	.437	1.000		
Pess	.396	.179	.232	.153	.320	.185	.322	.077	.398	.180	.109	.480	.402	.374	.319	.391	.226	.426	.328	1.000	
Sad	.143	.280	.249	.409	.286	.237	.329	.165	.323	.251	.098	.365	.251	.282	.368	.275	.520	.285	.350	.360	1.000

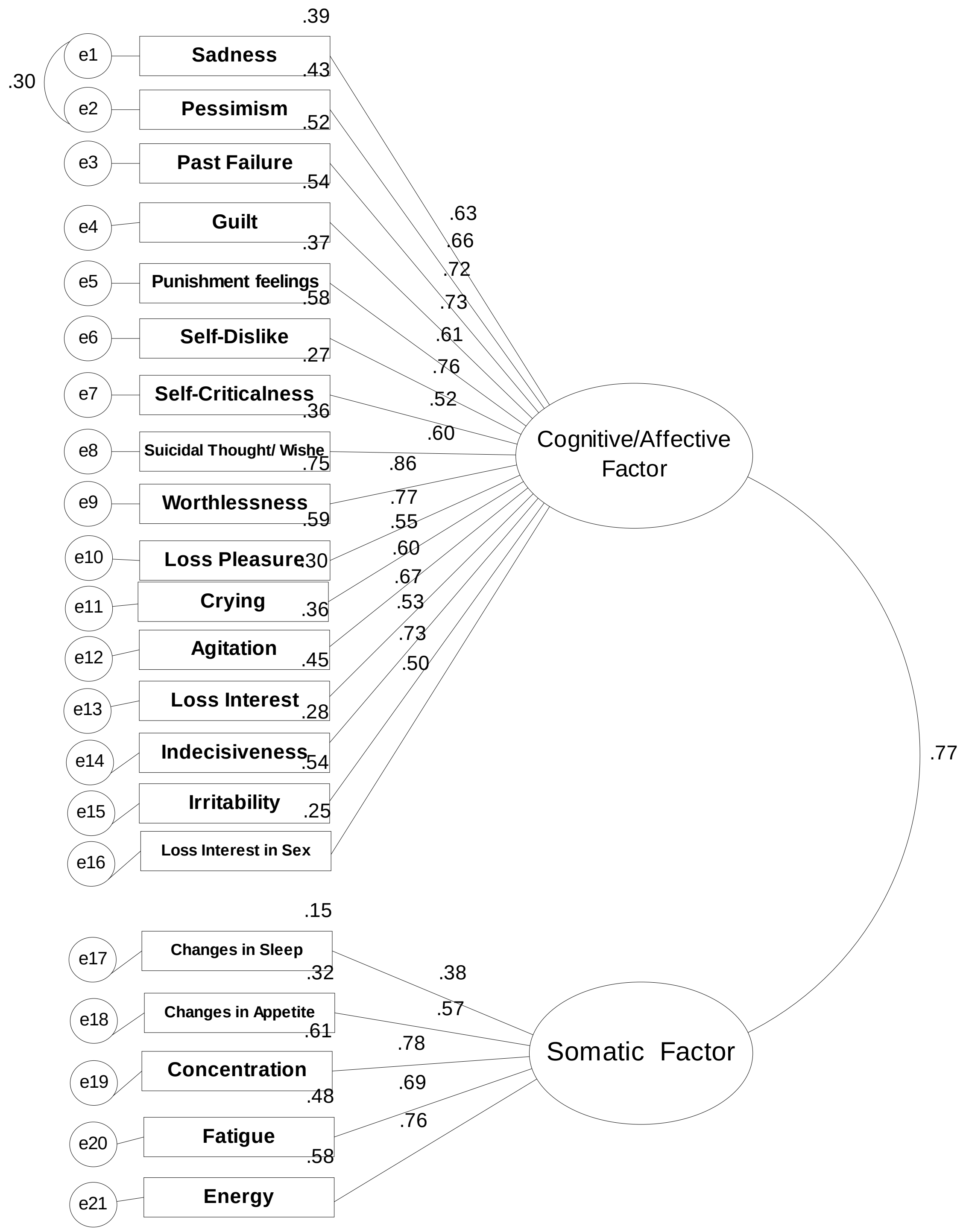
Residual Covariances

	Pleasure	Sex	Fatigue	Energy	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Appetite	Conc	Worth	Suicide	Critical	Dislike	Punish	Guilt	Failure	Pess	Sad
Pleasure	.221																				
Sex	.046	.113																			
Fatigue	.015	.050	.108																		
Energy	.064	.022	.060	.112																	
Cry	.004	.048	.012	.041	.160																
Agitate	.057	.004	.008	.017	.059	.078															
Interest	.079	.025	.014	.032	.061	.031	.137														
Indecisive	.065	.087	.057	.098	.075	.064	.062	.191													
Irritate	.065	.014	-.014	.016	.043	.037	.065	.049	.079												
Sleep	.033	.067	.012	.055	.086	.020	.015	.085	.017	.106											
Appetite	.028	.036	.014	.038	.067	.017	-.025	.055	.020	.016	.079										
Concentrate	.051	.031	.073	.085	.077	.032	.042	.079	.029	.050	.041	.115									
Worth	.168	.072	.042	.099	.105	.065	.079	.126	.081	.042	.060	.099	.210								
Suicide	.092	.000	.038	.051	.022	-.014	.042	.021	.046	.013	-.007	.029	.084	.089							
Critical	.055	-.004	.019	.047	.066	.038	.080	.058	.050	.012	.000	.076	.062	.041	.092						
Dislike	.125	.060	.071	.064	.073	.068	.035	.142	.075	.063	.052	.088	.119	.046	.058	.181					
Punish	.104	.102	.079	.036	.065	.010	.073	.054	.017	.027	.050	.055	.103	.052	.035	.072	.150				
Guilt	.082	.014	.010	.011	-.002	.022	.000	.035	.020	-.008	.006	.000	.052	.026	.022	.073	.021	.052			
Fail	.124	.032	-.010	.041	.030	.017	.002	.075	.048	.041	.029	.052	.153	.057	.010	.041	.039	.038	.167		
Pessimism	.129	.013	.031	.092	.058	.020	.077	.059	.062	.009	.032	.101	.127	.088	.093	.077	.107	.012	.073	.177	
Sad	.049	-.020	.019	.002	.031	.047	.033	.054	.023	.012	.011	.033	.038	.024	.018	.050	.017	.016	.025	.054	.056

Standardized Residual Covariances

	Pleasure	Sex	Fatigue	Energy	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	App	Conc.	Worth	Suicide	Critical	Dislike	Punish	Guilt	Fail	Pessi	Sad
Pleasure	13.017																				
Sex	2.467	2.870																			
Fatigue	.846	1.840	3.289																		
Energy	5.425	1.198	3.511	7.140																	
Cry	.290	2.261	.590	3.043	7.376																
Agitate	3.597	.151	.330	1.095	3.298	2.887															
Interest	6.259	1.278	.753	2.619	4.224	1.879	7.627														
Indecisi	4.286	3.743	2.702	6.716	4.402	3.322	3.997	7.058													
Irritate	4.863	.656	-.702	1.216	2.804	2.071	4.623	2.999	4.144												
Sleep	1.816	2.436	.474	3.154	4.225	.876	.822	3.813	.872	2.866											
Appetite	1.469	1.234	.502	2.054	3.122	.682	-1.267	2.392	.943	.564	2.034										
Conc	3.256	1.271	3.204	5.604	4.333	1.535	2.549	4.127	1.638	2.187	1.643	4.284									
Worth	14.333	3.985	2.544	8.683	7.921	4.332	6.515	8.639	6.344	2.460	3.325	6.639	13.310								
Suicide	8.062	-.028	2.361	4.595	1.735	-.995	3.548	1.484	3.817	.747	-.387	2.017	7.557	5.853							
Critical	3.302	-.170	.822	2.915	3.545	1.803	4.689	2.805	2.793	.483	-.008	3.646	3.788	2.608	2.878						
Dislike	11.288	3.545	4.558	5.978	5.802	4.811	3.092	10.208	6.289	3.860	3.043	6.242	10.805	4.338	3.755	12.605					
Punish	6.243	4.012	3.357	2.237	3.442	.448	4.234	2.606	.969	1.095	1.948	2.597	6.227	3.271	1.509	4.671	4.675				
Guilt	4.773	.512	.414	.676	-.121	.967	-.006	1.657	1.046	-.311	.238	-.011	2.881	1.591	.896	4.453	.848	1.577			
Fail	8.749	1.493	-.492	3.016	1.861	.943	.113	4.242	3.086	1.954	1.353	2.874	10.665	4.250	.519	3.121	1.968	1.750	7.238		
Pessi	9.518	.642	1.633	7.044	3.785	1.152	5.496	3.482	4.227	.451	1.528	5.870	9.485	6.853	4.932	6.128	5.647	.600	4.532	8.215	
Sad	3.733	-1.005	1.007	.186	2.092	2.735	2.421	3.265	1.582	.594	.548	1.935	2.757	1.877	.962	3.981	.908	.777	1.497	3.543	2.819

Appendix F - 2 FACTOR BDI COGNITIVE/AFFECTIVE FACTOR & SOMATIC FACTOR



The model is recursive.

Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

Observed, endogenous variables
Loss of Pleasure
Guilt Feelings
Punishment Feelings
Worthlessness
Suicidal Thoughts or wishes
Self-Critical
Self-Dislike
Sadness
Pessimism
Past Failure
Loss of interest in Sex
Concentration difficulty
Change in Appetite
Change in Sleep
Irritated
Indecisiveness
Loss of Interest
Agitation
Crying
Tiredness or Fatigue
Loss of Energy

Unobserved, exogenous variables
Cognitive_Affective

Somatic

Variable counts (including residuals)

Number of variables in your model:	44
Number of observed variables:	21
Number of unobserved variables:	23
Number of exogenous variables:	23
Number of endogenous variables:	21

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	23	0	0	0	0	23
Labeled	0	0	0	0	0	0
Unlabeled	19	2	23	0	0	44
Total	42	2	23	0	0	67

Computation of degrees of freedom

Number of distinct sample moments:	231
Number of distinct parameters to be estimated:	44
Degrees of freedom (231 - 44):	187

Result

Minimum was achieved

Chi-square = 215.477

Degrees of freedom = 187

Probability level = .075

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
BDI2factor	44	215.477	187	.075	1.152
Saturated model	231	.000	0		
Independence model	21	278.161	210	.001	1.325
Zero model	0	1039.500	231	.000	4.500

GFI

Model	GFI
BDI2factor	.793
Saturated model	1.000
Independence model	.732
Zero model	.000

Baseline Comparisons

Model	CFI
BDI2factor	.582
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
BDI2factor	.039	.000	.061	.770
Independence model	.057	.037	.075	.253

Regression Weights:

			Estimate	S.E.	C.R.	P
Guilt	<---	Cognitive_Affective	1.396	.322	4.333	***
Punish	<---	Cognitive_Affective	1.391	.504	2.761	.006
Worthlessness	<---	Cognitive_Affective	1.962	.577	3.401	***
Suicide	<---	Cognitive_Affective	.983	.360	2.732	.006
Critical	<---	Cognitive_Affective	1.124	.370	3.038	.002
Dislike	<---	Cognitive_Affective	1.460	.448	3.260	.001
Sad	<---	Cognitive_Affective	1.000			
Pessimism	<---	Cognitive_Affective	1.437	.493	2.915	.004
Failure	<---	Cognitive_Affective	1.417	.489	2.900	.004
Fatigue	<---	Somatic	.900	.195	4.620	***
Energy	<---	Somatic	.834	.180	4.636	***
Irritate	<---	Cognitive_Affective	1.198	.397	3.017	.003
Indecisive	<---	Cognitive_Affective	1.154	.476	2.423	.015
Lost Interest	<---	Cognitive_Affective	1.150	.475	2.425	.015
Cry	<---	Cognitive_Affective	.959	.471	2.036	.042
Agitate	<---	Cognitive_Affective	1.063	.354	3.002	.003
Sleep	<---	Somatic	.500	.206	2.427	.015
Appetite	<---	Somatic	.712	.200	3.556	***
Concentrate	<---	Somatic	1.000			
Pleasure	<---	Cognitive_Affective	1.689	.557	3.031	.002
Sex	<---	Cognitive_Affective	1.052	.516	2.038	.042

Standardized Regression Weights:

			Estimate
Guilt	<---	Cognitive_Affective	.732
Punish	<---	Cognitive_Affective	.609
Worthlessness	<---	Cognitive_Affective	.864
Suicide	<---	Cognitive_Affective	.598
Critical	<---	Cognitive_Affective	.520
Dislike	<---	Cognitive_Affective	.765
Sad	<---	Cognitive_Affective	.625
Pessimism	<---	Cognitive_Affective	.659
Failure	<---	Cognitive_Affective	.721
Fatigue	<---	Somatic	.693
Energy	<---	Somatic	.760
Irritate	<---	Cognitive_Affective	.733
Indecisive	<---	Cognitive_Affective	.531
Lost Interest	<---	Cognitive_Affective	.673
Cry	<---	Cognitive_Affective	.550
Agitate	<---	Cognitive_Affective	.602
Sleep	<---	Somatic	.384
Appetite	<---	Somatic	.567
Concentrate	<---	Somatic	.779
Pleasure	<---	Cognitive_Affective	.766
Sex	<---	Cognitive_Affective	.495

Covariances:

			Estimate	S.E.	C.R.	P
Cognitive_Affective	<-->	Somatic	.070	.025	2.750	.006
e1	<-->	e2	.034	.015	2.302	.021

Correlations:

			Estimate
Cognitive_Affective	<-->	Somatic	.769
e1	<-->	e2	.305

Variances:

	Estimate	S.E.	C.R.	P
Cognitive_Affective	.055	.028	1.940	.052
Somatic	.151	.050	2.999	.003
e1	.085	.016	5.375	***
e2	.147	.026	5.564	***
e3	.101	.023	4.451	***
e5	.093	.020	4.637	***
e6	.180	.033	5.514	***
e7	.083	.020	4.185	***
e8	.187	.031	6.051	***
e9	.095	.018	5.387	***
e10	.072	.016	4.551	***
e18	.098	.022	4.344	***
e17	.161	.031	5.139	***
e16	.218	.038	5.797	***
e15	.068	.015	4.585	***
e13	.087	.020	4.282	***
e12	.109	.021	5.186	***
e11	.116	.025	4.738	***
e21	.077	.019	4.145	***
e20	.132	.028	4.708	***
e4	.110	.024	4.672	***
e14	.186	.035	5.318	***
e19	.187	.035	5.358	***

Squared Multiple Correlations:

	Estimate
Sex	.245
Indecisive	.282
Pleas	.587
Fatigue	.481
Energy	.577
Cry	.302
Agitate	.362
Interest	.454
Irritate	.538
Sleep	.147
Appetite	.322
Concentrate	.607
Worth	.746
Suicide	.357
Critical	.271
Dislike	.585
Punish	.371
Guilt	.535
Fail	.520
Pess	.435
Sad	.391

Sample Covariances

	Sex	Indecis	Pleas	Fatigue	Energy	Cry	Agitate	Interest	Irritate	Sleep	Appetite	Conc	Worth	Suicide	Critical	Dislike	Punish	Guilt	Fail	Pess	Sad
Sex	.390																				
Indecisive	.106	.380																			
Pleas	.088	.075	.340																		
Fatigue	.143	.079	.062	.340																	
Energy	.070	.109	.088	.113	.223																
Cry	.101	.088	.031	.071	.071	.312															
Agitate	.086	.083	.098	.100	.064	.111	.268														
Interest	.080	.075	.107	.076	.064	.096	.086	.263													
Irritate	.091	.067	.104	.073	.061	.092	.114	.117	.214												
Sleep	.124	.098	.061	.076	.088	.122	.077	.053	.070	.365											
Appetite	.132	.077	.076	.121	.093	.128	.112	.039	.110	.081	.353										
Concentrate	.111	.097	.091	.163	.132	.128	.111	.095	.104	.105	.133	.303									
Worth	.108	.135	.186	.083	.120	.128	.102	.104	.115	.068	.102	.135	.321								
Suicide	.017	.025	.101	.057	.061	.033	.003	.053	.063	.025	.013	.046	.104	.196							
Critical	.031	.066	.073	.059	.067	.089	.073	.104	.083	.036	.041	.111	.102	.060	.316						
Dislike	.085	.148	.138	.099	.079	.089	.093	.052	.098	.080	.080	.112	.147	.059	.085	.282					
Punish	.141	.063	.124	.123	.059	.090	.048	.099	.054	.054	.095	.093	.147	.073	.078	.103	.376				
Guilt	.086	.052	.118	.091	.053	.044	.093	.048	.087	.042	.090	.069	.134	.065	.101	.129	.109	.284			
Fail	.076	.085	.146	.039	.067	.058	.060	.031	.088	.071	.080	.094	.203	.081	.058	.075	.092	.136	.329		
Pess	.039	.065	.142	.060	.107	.074	.045	.094	.086	.027	.061	.126	.156	.102	.121	.097	.138	.069	.108	.329	
Sad	.027	.065	.073	.072	.030	.061	.094	.065	.067	.044	.066	.079	.091	.049	.070	.086	.075	.123	.089	.091	.195

Sample Correlations

	Sex	Indecis	Pleas	Fatigue	Energy	Cry	Agitate	Interest	Irritate	Sleep	Appetite	Conc	Worth	Suicide	Critical	Dislike	Punish	Guilt	Fail	Pess	Sad
Sex	1.000																				
Indecisive	.275	1.000																			
Pleas	.243	.208	1.000																		
Fatigue	.392	.219	.182	1.000																	
Energy	.236	.375	.321	.411	1.000																
Cry	.289	.254	.094	.217	.270	1.000															
Agitate	.265	.259	.326	.331	.263	.385	1.000														
Interest	.250	.238	.359	.253	.265	.334	.323	1.000													
Irritate	.315	.233	.386	.270	.277	.355	.475	.492	1.000												
Sleep	.328	.264	.174	.215	.307	.361	.245	.172	.251	1.000											
Appetite	.355	.211	.220	.350	.332	.385	.364	.127	.398	.226	1.000										
Concentrate	.322	.287	.285	.507	.506	.416	.389	.336	.407	.315	.406	1.000									
Worth	.306	.386	.565	.253	.447	.406	.346	.356	.439	.197	.304	.432	1.000								
Suicide	.061	.093	.390	.222	.290	.135	.013	.235	.306	.092	.051	.188	.415	1.000							
Critical	.089	.192	.222	.180	.253	.283	.251	.360	.318	.106	.122	.357	.321	.242	1.000						
Dislike	.257	.452	.446	.320	.313	.299	.337	.191	.401	.250	.255	.383	.490	.252	.285	1.000					
Punish	.369	.168	.346	.343	.202	.262	.152	.314	.190	.145	.260	.274	.424	.268	.226	.315	1.000				
Guilt	.257	.158	.381	.293	.210	.146	.338	.175	.354	.130	.283	.236	.444	.275	.338	.454	.332	1.000			
Fail	.212	.240	.437	.117	.246	.180	.203	.104	.333	.204	.234	.298	.624	.318	.180	.247	.262	.444	1.000		
Pess	.109	.185	.426	.179	.396	.232	.153	.320	.322	.077	.180	.398	.480	.402	.374	.319	.391	.226	.328	1.000	
Sad	.098	.237	.285	.280	.143	.249	.409	.286	.329	.165	.251	.323	.365	.251	.282	.368	.275	.520	.350	.360	1.000

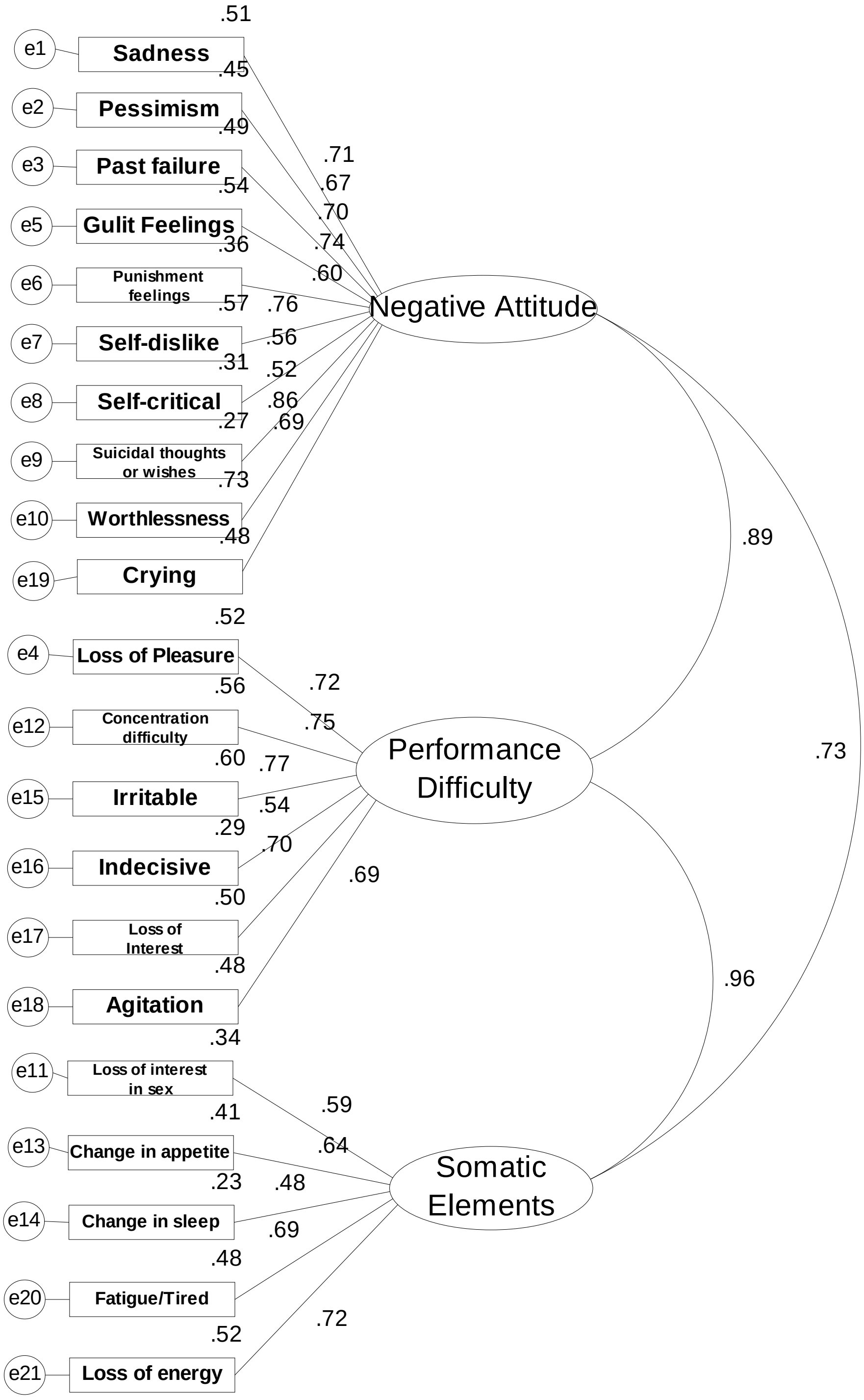
Residual Covariances

	Sex	Indecis	Pleas	Fatigue	Energy	Cry	Agitate	Interest	Irritate	Sleep	Appetite	Conc	Worth	Suicide	Critical	Dislike	Punish	Guilt	Fail	Pess	Sad
Sex	.143																				
Indecis	.039	.122																			
Pleas	-.009	-.032	.074																		
Fatigue	.076	.006	-.044	.086																	
Energy	.008	.042	-.010	.000	.041																
Cry	.045	.027	-.058	.010	.015	.145															
Agitate	.025	.016	.000	.033	.002	.056	.098														
Interest	.014	.002	.001	.003	-.003	.035	.019	.103													
Irritate	.022	-.009	-.007	-.002	-.009	.029	.044	.041	.068												
Sleep	.087	.058	.002	.008	.025	.088	.039	.013	.028	.109											
Appetite	.079	.020	-.008	.024	.003	.080	.059	-.019	.050	.028	.115										
Conc	.037	.017	-.027	.027	.006	.061	.037	.015	.020	.029	.025	.055									
Worth	-.005	.011	.005	-.040	.005	.025	-.013	-.020	-.014	-.001	.005	-.002	.038								
Suicide	-.040	-.037	.010	-.005	.003	-.018	-.054	-.009	-.002	-.010	-.036	-.023	-.002	.048							
Critical	-.034	-.005	-.031	-.012	.002	.030	.008	.033	.009	-.003	-.015	.032	-.019	.000	.060						
Dislike	.001	.056	.003	.007	-.007	.012	.008	-.040	.003	.029	.008	.010	-.010	-.019	-.005	.082					
Punish	.061	-.025	-.005	.035	-.023	.017	-.033	.011	-.037	.005	.026	-.005	-.002	-.002	-.008	-.009	.091				
Guilt	.005	-.036	-.011	.003	-.028	-.030	.012	-.040	-.004	-.007	.020	-.028	-.016	-.010	.015	.017	.002	.085			
Fail	-.006	-.005	.015	-.050	-.016	-.017	-.022	-.059	-.005	.021	.009	-.005	.050	.004	-.029	-.038	-.016	.027	.118		
Pess	-.044	-.026	.009	-.031	.023	-.001	-.038	.003	-.009	-.024	-.010	.025	.001	.025	.032	-.018	.028	-.041	-.004	.068	
Sad	-.031	.001	-.019	.009	-.029	.009	.036	.002	.002	.009	.016	.009	-.016	-.005	.008	.006	-.002	.046	.011	-.022	.055

Standardized Residual Covariances

	Sex	Indecis	Pleas	Fatigue	Energy	Cry	Agitate	Interest	Irritate	Sleep	Appetite	Conc	Worth	Suicide	Critical	Dislike	Punish	Guilt	Fail	Pess	Sad
Sex	4.066																				
Indecisive	1.493	3.305																			
Pleas	-.330	-1.126	1.946																		
Fatigue	2.932	.231	-1.571	2.369																	
Energy	.375	1.840	-.416	.002	1.599																
Cry	2.149	1.243	-2.522	.482	.835	6.125															
Agitate	1.140	.707	-.001	1.500	.129	3.106	4.020														
Interest	.655	.114	.032	.151	-.160	2.021	1.052	4.543													
Irritate	1.077	-.441	-.293	-.120	-.516	1.711	2.538	2.404	3.269												
Sleep	3.406	2.216	.089	.299	1.093	4.200	1.851	.624	1.421	2.983											
Appetite	3.184	.775	-.292	.921	.152	3.886	2.824	-.908	2.535	1.085	3.412										
Concentrate	1.427	.625	-.937	.939	.233	2.820	1.668	.675	.948	1.113	.950	1.545									
Worth	-.165	.358	.149	-1.350	.208	1.051	-.506	-.814	-.560	-.036	.166	-.083	.949								
Suicide	-1.982	-1.785	.440	-.228	.186	-1.095	-3.195	-.522	-.109	-.495	-1.821	-1.119	-.076	2.266							
Critical	-1.291	-.173	-1.112	-.446	.067	1.381	.352	1.525	.434	-.122	-.597	1.204	-.626	-.023	1.655						
Dislike	.038	2.257	.107	.297	-.315	.605	.382	-1.979	.133	1.252	.334	.397	-.337	-1.025	-.197	2.896					
Punish	2.189	-.854	-.164	1.230	-.926	.718	-1.378	.481	-1.659	.185	.943	-.164	-.062	-.097	-.274	-.327	2.239				
Guilt	.218	-1.486	-.403	.139	-1.363	-1.507	.594	-1.999	-.218	-.303	.885	-1.154	-.557	-.538	.627	.742	.087	3.004			
Fail	-.235	-.180	.545	-2.010	-.752	-.832	-1.061	-2.854	-.237	.882	.394	-.203	1.745	.227	-1.166	-1.620	-.584	1.176	3.917		
Pess	-1.639	-.923	.313	-1.116	.998	-.048	-1.682	.156	-.405	-.893	-.396	.910	.044	1.163	1.172	-.701	.954	-1.601	-.145	1.849	
Sad	-1.569	.065	-.892	.461	-1.668	.549	2.137	.108	.100	.471	.852	.429	-.704	-.301	.420	.340	-.072	2.492	.582	-.975	2.769

Appendix G – 3-FACTOR CFA Model of BECK DEPRESSION INVENTORY



The model is recursive.

Sample size = 100

Your model contains the following variables (residuals not shown)

Observed, endogenous variables

Loss of Pleasure
Guilt Feelings
Punishment Feelings
Worthlessness
Suicidal Thoughts or wishes
Self-Critical
Self-Dislike
Sadness
Pessimism
Past Failure
Loss of interest in Sex
Concentration difficulty
Change in Appetite
Change in Sleep
Irritated
Indecisiveness
Loss of Interest
Agitation
Crying
Tiredness or Fatigue
Loss of Energy

Unobserved, exogenous variables

negative attiude

performance difficulty

somatic elements

Variable counts

Number of variables in your model: 45
Number of observed variables: 21
Number of unobserved variables: 24
Number of exogenous variables: 24
Number of endogenous variables: 21

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	24	0	0	0	0	24
Labeled	0	0	0	0	0	0
Unlabeled	18	3	24	0	0	45
Total	42	3	24	0	0	69

Computation of degrees of freedom

Number of distinct sample moments: 231
Number of distinct parameters to be estimated: 45

Degrees of freedom (231 - 45): 186

Result

Minimum was achieved
Chi-square = 216.952
Degrees of freedom = 186
Probability level = .060

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
BDI1factor model	45	216.952	186	.060	1.166
Saturated model	231	.000	0		
Independence model	21	278.161	210	.001	1.325
Zero model	0	1039.500	231	.000	4.500

GFI

Model	GFI
BDI1factor model	.791
Saturated model	1.000
Independence model	.732
Zero model	.000

Baseline Comparisons

Model	CFI
BDI1factor model	.546
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
BDI1factor model	.041	.000	.062	.732
Independence model	.057	.037	.075	.253

Regression Weights:

			Estimate	S.E.	C.R.	P
Pessimism	<---	negative attiude	.741	.126	5.899	***
Failure	<---	negative attiude	.723	.109	6.644	***
Guilt	<---	negative attiude	.744	.123	6.053	***
Loss Pleasure	<---	performance difficulty	1.215	.319	3.808	***
Punish	<---	negative attiude	.722	.137	5.283	***
Self-Dislike	<---	negative attiude	.750	.118	6.348	***
Self-Critical	<---	negative attiude	.661	.136	4.859	***
Suicide	<---	negative attiude	.428	.096	4.453	***
Worthlessness	<---	negative attiude	1.000			
Agitation	<---	performance difficulty	1.154	.287	4.015	***
Loss Interest	<---	performance difficulty	1.069	.275	3.885	***
Indecisive	<---	performance difficulty	1.000			
Irritability	<---	performance difficulty	1.144	.282	4.053	***
Concentrate	<---	performance difficulty	1.376	.328	4.196	***
Energy	<---	somatic elements	1.000			
Fatigue	<---	somatic elements	1.185	.228	5.197	***
Appetite	<---	somatic elements	1.145	.241	4.753	***
Sex	<---	somatic elements	1.094	.259	4.230	***
Sleep	<---	somatic elements	.878	.225	3.899	***
Crying	<---	negative attiude	.734	.129	5.674	***
Sadness	<---	negative attiude	.628	.108	5.801	***

Standardized Regression Weights

			Estimate
Pessimism	<---	negative attiude	.670
Failure	<---	negative attiude	.702
Guilt	<---	negative attiude	.735
Loss Pleasure	<---	performance difficulty	.722
Punish	<---	negative attiude	.598
Self-Dislike	<---	negative attiude	.758
Self-Critical	<---	negative attiude	.557
Suicide	<---	negative attiude	.515
Worthlessness	<---	negative attiude	.857
Agitation	<---	performance difficulty	.691
Loss Interest	<---	performance difficulty	.704
Indecisive	<---	performance difficulty	.537
Irritability	<---	performance difficulty	.775
Concentrate	<---	performance difficulty	.750
Energy	<---	somatic elements	.722
Fatigue	<---	somatic elements	.691
Appetite	<---	somatic elements	.639
Sex	<---	somatic elements	.587
Sleep	<---	somatic elements	.479
Crying	<---	negative attiude	.691
Sadness	<---	negative attiude	.714

Covariances:

			Estimate	S.E.	C.R.	P
performance difficulty	<-->	somatic elements	.077	.023	3.302	***
negative attiude	<-->	performance difficulty	.107	.030	3.543	***
negative attiude	<-->	somatic elements	.093	.024	3.932	***

Correlations:

			Estimate
performance difficulty	<-->	somatic elements	.960
negative attiude	<-->	performance difficulty	.890
negative attiude	<-->	somatic elements	.734

Variances:

	Estimate	S.E.	C.R.	P
negative attiude	.191	.042	4.499	***
performance difficulty	.076	.033	2.281	.023
somatic elements	.085	.026	3.231	.001
e1	.072	.015	4.752	***
e2	.129	.025	5.151	***
e3	.103	.023	4.522	***
e4	.103	.024	4.326	***
e5	.090	.020	4.509	***
e6	.179	.033	5.495	***
e7	.080	.020	4.031	***
e8	.186	.031	5.995	***
e9	.097	.018	5.518	***
e10	.069	.016	4.368	***
e11	.193	.035	5.494	***
e13	.162	.031	5.133	***
e12	.112	.022	5.015	***
e14	.219	.038	5.818	***
e15	.066	.015	4.416	***
e16	.186	.035	5.256	***
e17	.088	.020	4.304	***
e18	.110	.021	5.232	***
e19	.113	.025	4.485	***
e20	.130	.027	4.776	***
e21	.078	.017	4.465	***

Squared Multiple Correlations:

	Estimate
Energy	.521
Fatigue	.478
Cry	.477
Agitate	.477
Interest	.496
Indecisive	.289
Irritate	.600
Sleep	.230
Conc	.562
Appetite	.408
Sex	.345
Worth	.735
Suicide	.266
Critical	.310
Dislike	.575
Punish	.358
Guilt	.540
Pleas	.521
Fail	.493
Pess	.449
Sad	.509

Sample Moments

Sample Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	.223																				
Fatigue	.113	.340																			
Cry	.071	.071	.312																		
Agitate	.064	.100	.111	.268																	
Interest	.064	.076	.096	.086	.263																
Indecisive	.109	.079	.088	.083	.075	.380															
Irritate	.061	.073	.092	.114	.117	.067	.214														
Sleep	.088	.076	.122	.077	.053	.098	.070	.365													
Conc	.132	.163	.128	.111	.095	.097	.104	.105	.303												
Appetite	.093	.121	.128	.112	.039	.077	.110	.081	.133	.353											
Sex	.070	.143	.101	.086	.080	.106	.091	.124	.111	.132	.390										
Worth	.120	.083	.128	.102	.104	.135	.115	.068	.135	.102	.108	.321									
Suicide	.061	.057	.033	.003	.053	.025	.063	.025	.046	.013	.017	.104	.196								
Critical	.067	.059	.089	.073	.104	.066	.083	.036	.111	.041	.031	.102	.060	.316							
Dislike	.079	.099	.089	.093	.052	.148	.098	.080	.112	.080	.085	.147	.059	.085	.282						
Punish	.059	.123	.090	.048	.099	.063	.054	.054	.093	.095	.141	.147	.073	.078	.103	.376					
Guilt	.053	.091	.044	.093	.048	.052	.087	.042	.069	.090	.086	.134	.065	.101	.129	.109	.284				
Pleas	.088	.062	.031	.098	.107	.075	.104	.061	.091	.076	.088	.186	.101	.073	.138	.124	.118	.340			
Fail	.067	.039	.058	.060	.031	.085	.088	.071	.094	.080	.076	.203	.081	.058	.075	.092	.136	.146	.329		
Pess	.107	.060	.074	.045	.094	.065	.086	.027	.126	.061	.039	.156	.102	.121	.097	.138	.069	.142	.108	.329	
Sad	.030	.072	.061	.094	.065	.065	.067	.044	.079	.066	.027	.091	.049	.070	.086	.075	.123	.073	.089	.091	.195

Sample Correlations

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	1.000																				
Fatigue	.411	1.000																			
Cry	.270	.217	1.000																		
Agitate	.263	.331	.385	1.000																	
Interest	.265	.253	.334	.323	1.000																
Indecisive	.375	.219	.254	.259	.238	1.000															
Irritate	.277	.270	.355	.475	.492	.233	1.000														
Sleep	.307	.215	.361	.245	.172	.264	.251	1.000													
Conc	.506	.507	.416	.389	.336	.287	.407	.315	1.000												
Appetite	.332	.350	.385	.364	.127	.211	.398	.226	.406	1.000											
Sex	.236	.392	.289	.265	.250	.275	.315	.328	.322	.355	1.000										
Worth	.447	.253	.406	.346	.356	.386	.439	.197	.432	.304	.306	1.000									
Suicide	.290	.222	.135	.013	.235	.093	.306	.092	.188	.051	.061	.415	1.000								
Critical	.253	.180	.283	.251	.360	.192	.318	.106	.357	.122	.089	.321	.242	1.000							
Dislike	.313	.320	.299	.337	.191	.452	.401	.250	.383	.255	.257	.490	.252	.285	1.000						
Punish	.202	.343	.262	.152	.314	.168	.190	.145	.274	.260	.369	.424	.268	.226	.315	1.000					
Guilt	.210	.293	.146	.338	.175	.158	.354	.130	.236	.283	.257	.444	.275	.338	.454	.332	1.000				
Pleas	.321	.182	.094	.326	.359	.208	.386	.174	.285	.220	.243	.565	.390	.222	.446	.346	.381	1.000			
Fail	.246	.117	.180	.203	.104	.240	.333	.204	.298	.234	.212	.624	.318	.180	.247	.262	.444	.437	1.000		
Pess	.396	.179	.232	.153	.320	.185	.322	.077	.398	.180	.109	.480	.402	.374	.319	.391	.226	.426	.328	1.000	
Sad	.143	.280	.249	.409	.286	.237	.329	.165	.323	.251	.098	.365	.251	.282	.368	.275	.520	.285	.350	.360	1.000

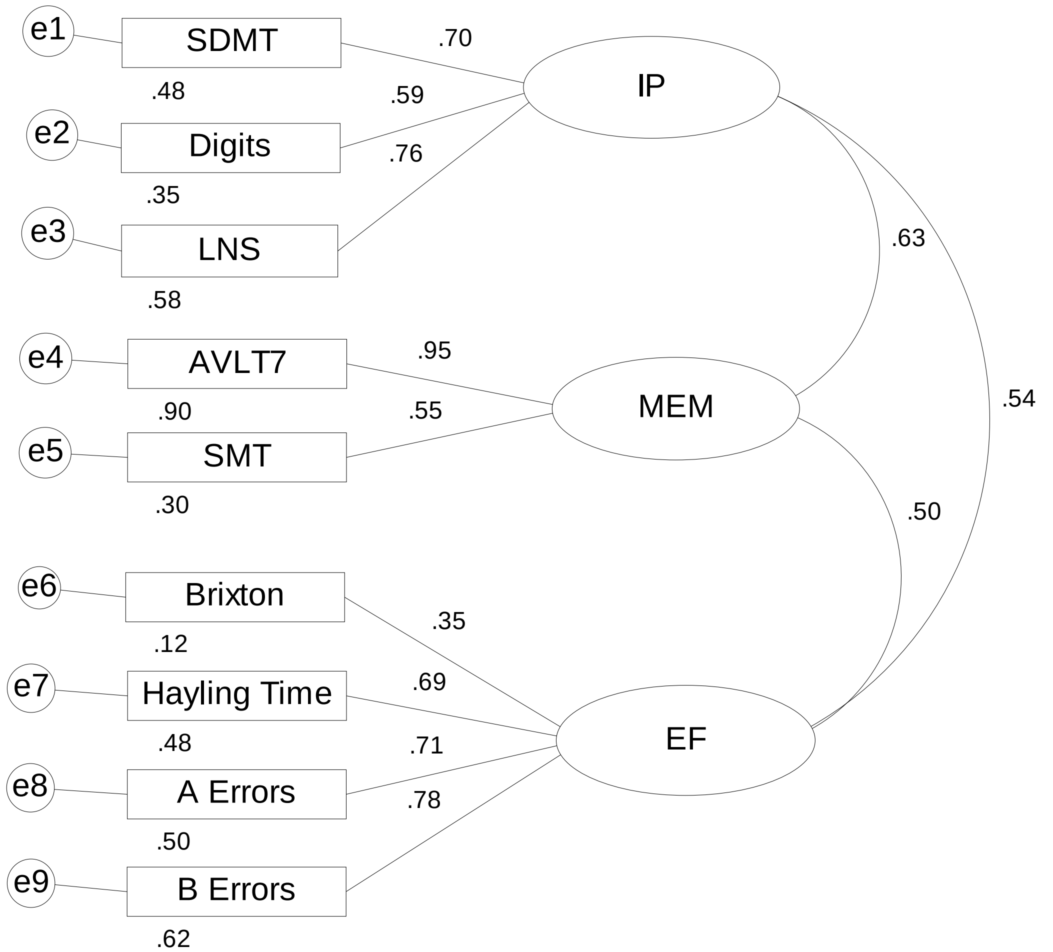
Residual Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	.061																				
Fatigue	.013	.090																			
Cry	.003	-.011	.097																		
Agitate	-.024	-.005	.021	.057																	
Interest	-.018	-.022	.012	-.007	.089																
Indecisive	.032	-.012	.009	-.004	-.006	.119															
Irritate	-.027	-.031	.002	.014	.024	-.020	.049														
Sleep	.013	-.013	.062	-.001	-.019	.031	-.007	.080													
Conc	.026	.037	.020	-.009	-.016	-.007	-.015	.012	.048												
Appetite	-.004	.006	.049	.010	-.055	-.011	.009	-.004	.012	.080											
Sex	-.023	.033	.026	-.011	-.010	.022	-.005	.042	-.005	.025	.096										
Worth	.026	-.027	-.012	-.022	-.011	.028	-.007	-.015	-.013	-.005	.006	.061									
Suicide	.021	.010	-.027	-.050	.004	-.020	.010	-.011	-.017	-.032	-.027	.022	.064								
Critical	.005	-.014	-.004	-.008	.028	-.004	.002	-.018	.013	-.030	-.036	-.024	.006	.047							
Dislike	.008	.016	-.016	.000	-.034	.068	.007	.019	.001	.000	.008	.004	-.002	-.010	.095						
Punish	-.009	.043	-.012	-.041	.016	-.014	-.034	-.006	-.014	.018	.067	.009	.014	-.013	-.001	.098					
Guilt	-.016	.009	-.061	.001	-.037	-.028	-.004	-.019	-.040	.010	.010	-.008	.004	.007	.022	.006	.089				
Pleas	-.005	-.049	-.065	-.008	.009	-.017	-.001	-.021	-.035	-.031	-.014	.056	.045	-.013	.040	.030	.022	.125			
Fail	-.001	-.041	-.044	-.029	-.052	.008	.000	.011	-.012	.003	.002	.065	.022	-.033	-.028	-.008	.033	.052	.126		
Pess	.038	-.022	-.029	-.046	.009	-.014	-.005	-.034	.016	-.018	-.037	.014	.042	.027	-.009	.035	-.036	.046	.006	.095	
Sad	-.029	.003	-.027	.016	-.007	-.003	-.010	-.007	-.014	-.001	-.037	-.028	-.002	-.009	-.004	-.012	.033	-.008	.002	.002	.048

Standardized Residual Covariances

	Energy	Fatigue	Cry	Agitate	Interest	Indecisive	Irritate	Sleep	Conc	Appetite	Sex	Worth	Suicide	Critical	Dislike	Punish	Guilt	Pleas	Fail	Pess	Sad
Energy	2.619																				
Fatigue	.565	2.546																			
Cry	.136	-.426	3.151																		
Agitate	-1.178	-.209	.891	1.903																	
Interest	-.957	-.941	.561	-.349	3.583																
Indecisive	1.467	-.450	.366	-.174	-.245	3.185															
Irritate	-1.461	-1.368	.100	.654	1.246	-.885	2.099														
Sleep	.570	-.445	2.406	-.052	-.810	1.090	-.310	1.969													
Conc	1.122	1.322	.764	-.347	-.679	-.238	-.650	.416	1.335												
Appetite	-.172	.212	1.925	.400	-2.318	-.377	.374	-.139	.403	2.076											
Sex	-.968	1.108	.975	-.419	-.401	.744	-.219	1.396	-.171	.838	2.282										
Worth	1.158	-.978	-.425	-.820	-.445	.981	-.297	-.509	-.420	-.160	.206	1.642									
Suicide	1.349	.519	-1.479	-2.836	.268	-1.064	.659	-.538	-.884	-1.649	-1.323	1.086	3.417								
Critical	.246	-.527	-.143	-.330	1.223	-.151	.087	-.639	.476	-1.061	-1.252	-.812	.307	1.231							
Dislike	.447	.690	-.721	.008	-1.678	2.862	.333	.775	.058	.007	.338	.148	-.131	-.391	3.559						
Punish	-.395	1.536	-.435	-1.570	.686	-.490	-1.478	-.194	-.478	.610	2.266	.311	.678	-.459	-.039	2.465					
Guilt	-.857	.371	-2.625	.063	-1.815	-1.150	-.185	-.790	-1.606	.417	.378	-.297	.239	.290	.998	.229	3.182				
Pleas	-.238	-1.894	-2.732	-.324	.411	-.672	-.044	-.789	-1.312	-1.152	-.511	2.085	2.527	-.516	1.807	1.132	.954	4.112			
Fail	-.049	-1.710	-1.875	-1.272	-2.517	.319	-.009	.454	-.492	.104	.082	2.404	1.237	-1.311	-1.283	-.290	1.465	2.266	4.392		
Pess	1.828	-.865	-1.183	-1.910	.426	-.530	-.237	-1.283	.613	-.675	-1.341	.501	2.223	1.010	-.391	1.283	-1.508	1.878	.228	2.863	
Sad	-1.730	.130	-1.327	.839	-.394	-.123	-.548	-.350	-.641	-.052	-1.697	-1.231	-.143	-.427	-.191	-.540	1.723	-.418	.105	.115	2.271

APPENDIX H - A PRIORI COGNITIVE MODEL



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables

SDMT	Symbol Digit Modalities test
Digits	Digit Span task
AVLT7	Delayed recall AVLT
SMTrecall	Delayed recall Sentence Memory Test
HAYTime	Hayling tests Time b *minus* Time A
ErrA	A Type Errors Hayling test
ErrB	B Type Errors Hayling tests
Brixton	Brixton test errors
LNS	letter Number Sequencing tests

Unobserved, exogenous variables

IP	Information processing
MEM	Memory
EF	Executive Function

Variable counts (including residuals)

Number of variables in your model:	21
Number of observed variables:	9
Number of unobserved variables:	12
Number of exogenous variables:	12
Number of endogenous variables:	9

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	12	0	0	0	0	12
Labeled	0	0	0	0	0	0
Unlabeled	6	3	12	0	0	21
Total	18	3	12	0	0	33

Computation of degrees of freedom

Number of distinct sample moments:	45
Number of distinct parameters to be estimated:	21
Degrees of freedom (45 - 21):	24

Result

Minimum was achieved
Chi-square = 33.772
Degrees of freedom = 24
Probability level = .089

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	21	33.772	24	.089	1.407
Saturated model	45	.000	0		
Independence model	9	89.223	36	.000	2.478
Zero model	0	445.500	45	.000	9.900

GFI

Model	GFI
Default model	.924
Saturated model	1.000
Independence model	.800
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.816
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.064	.000	.111	.300
Independence model	.122	.091	.154	.000

Regression Weights:

			Estimate	S.E.	C.R.	P
SDMT	<---	IP	1.000			
Digits	<---	IP	.158	.039	4.085	***
AVLT7	<---	MEM	1.000			
SMTrecall	<---	MEM	.630	.179	3.531	***
HAYTime	<---	EF	1.000			
ErrA	<---	EF	.028	.005	5.240	***
ErrB	<---	EF	.054	.010	5.544	***
Brixton	<---	EF	.058	.021	2.769	.006
LNS	<---	IP	.237	.049	4.802	***

Standardized Regression Weights:

			Estimate
SDMT	<---	IP	.695
Digits	<---	IP	.589
AVLT7	<---	MEM	.949
SMTrecall	<---	MEM	.550
HAYTime	<---	EF	.694
ErrA	<---	EF	.709
ErrB	<---	EF	.785
Brixton	<---	EF	.351
LNS	<---	IP	.763

Covariances:

			Estimate	S.E.	C.R.	P
F1	<-->	F2	15.491	4.416	3.508	***
F2	<-->	F3	53.700	17.146	3.132	.002
F1	<-->	F3	140.246	51.528	2.722	.006

Correlations:

			Estimate
F1	<-->	F2	.632
F2	<-->	F3	.505
F1	<-->	F3	.539

Variances:

	Estimate	S.E.	C.R.	P
F1	59.935	19.763	3.033	.002
F2	10.022	2.896	3.461	***
F3	1129.420	356.787	3.166	.002
e1	64.134	14.458	4.436	***
e2	2.800	.527	5.311	***
e4	1.118	2.304	.485	.628
e5	9.208	1.685	5.464	***
e7	1217.204	244.584	4.977	***
e8	.852	.180	4.745	***
e9	2.062	.591	3.492	***
e6	26.820	4.363	6.147	***
e3	2.400	.662	3.627	***

Squared Multiple Correlations:

	Estimate
LNS	.583
Brixton	.123
ErrB	.616
ErrA	.503
HAYTime	.481
SMTrecall	.302
AVLT7	.900
Digits	.347
SDMT	.483

Sample Moments

Sample Covariances

	LNS	Brixton	ErrB	ErrA	HAYTime	SMTrecall	AVLT7	Digits	SDMT
LNS	6.545								
Brixton	3.141	38.735							
ErrB	.751	3.402	6.054						
ErrA	.870	2.092	1.759	1.905					
HAYTime	13.867	99.076	77.538	35.509	2838.428				
SMTrecall	2.690	6.441	.917	1.889	47.646	16.462			
AVLT7	4.617	8.723	2.068	1.949	55.838	8.511	13.608		
Digits	2.550	1.372	.990	.460	15.780	1.979	2.170	4.540	
SDMT	14.555	29.043	8.959	6.160	254.354	15.103	22.784	7.570	152.365

Sample Correlations

	LNS	Brixton	ErrB	ErrA	HAYTime	SMTrecall	AVLT7	Digits	SDMT
LNS	1.000								
Brixton	.197	1.000							
ErrB	.119	.222	1.000						
ErrA	.246	.244	.518	1.000					
HAYTime	.102	.299	.591	.483	1.000				
SMTrecall	.259	.255	.092	.337	.220	1.000			
AVLT7	.489	.380	.228	.383	.284	.569	1.000		
Digits	.468	.103	.189	.156	.139	.229	.276	1.000	
SDMT	.461	.378	.295	.362	.387	.302	.500	.288	1.000

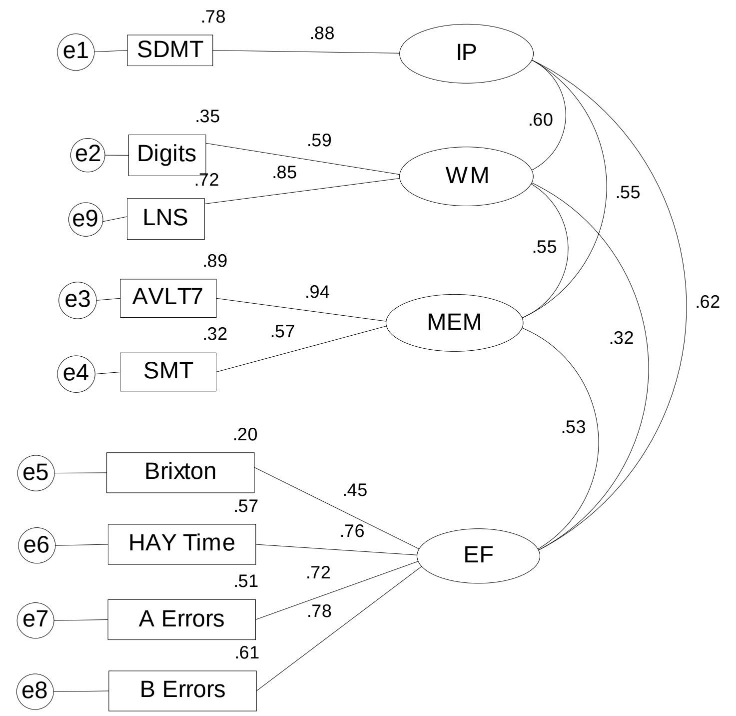
Matrices

Residual Covariances

	LNS	Brixton	ErrB	ErrA	HAYTime	SMTrecall	AVLT7	Digits	SDMT
LNS	.791								
Brixton	1.226	8.153							
ErrB	-1.043	-.123	.689						
ErrA	-.046	.293	.072	.192					
HAYTime	-19.310	33.890	16.458	4.326	491.804				
SMTrecall	.380	4.487	-.914	.955	13.790	3.270			
AVLT7	.952	5.623	-.836	.466	2.137	2.193	2.468		
Digits	.315	.096	-.206	-.150	-6.327	.439	-.272	.250	
SDMT	.377	20.949	1.374	2.288	114.107	5.337	7.293	-1.877	28.296

Standardized Residual Covariances

	LNS	Brixton	ErrB	ErrA	HAYTime	SMTrecall	AVLT7	Digits	SDMT
LNS	.967								
Brixton	.910	1.876							
ErrB	-1.777	-.092	.903						
ErrA	-.140	.391	.208	.789					
HAYTime	-1.590	1.223	1.282	.609	1.475				
SMTrecall	.419	2.212	-1.057	1.961	.766	1.744			
AVLT7	1.076	2.990	-1.007	1.005	.125	1.596	1.559		
Digits	.576	.083	-.414	-.538	-.613	.569	-.369	.411	
SDMT	.124	3.355	.508	1.509	2.036	1.276	1.802	-.749	1.605



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables

SDMT	Symbol Digit Modalities test
Digits	Digit Span task
AVLT7	Delayed recall AVLT
SMTrecall	Delayed recall Sentence Memory Test
HAYTime	Hayling tests Time b *minus* Time A
ErrA	A Type Errors Hayling test
ErrB	B Type Errors Hayling tests
Brixton	Brixton test errors
LNS	letter Number Sequencing tests

Unobserved, exogenous variables

IP	Information processing
WM	Working Memory
MEM	Memory
EF	Executive Function

Variable counts

Number of variables in your model:	22
Number of observed variables:	9
Number of unobserved variables:	13
Number of exogenous variables:	13
Number of endogenous variables:	9

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	13	0	1	0	0	14
Labeled	0	0	0	0	0	0
Unlabeled	5	6	12	0	0	23
Total	18	6	13	0	0	37

Computation of degrees of freedom

Number of distinct sample moments:	45
Number of distinct parameters to be estimated:	23
Degrees of freedom (45 - 23):	22

Result (Default model)

Minimum was achieved
Chi-square = 24.370
Degrees of freedom = 22
Probability level = .328

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	23	24.370	22	.328	1.108
Saturated model	45	.000	0		
Independence model	9	89.223	36	.000	2.478
Zero model	0	445.500	45	.000	9.900

GFI

Model	GFI
Default model	.945
Saturated model	1.000
Independence model	.800
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.955
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.033	.000	.093	.617
Independence model	.122	.091	.154	.000

Regression Weights:

			Estimate	S.E.	C.R.	P
AVLT7	<---	DR	1.000			
SMTrecall	<---	DR	.641	.166	3.869	***
HAYTime	<---	EF	1.000			
ErrA	<---	EF	.024	.004	5.947	***
ErrB	<---	EF	.047	.007	6.465	***
SDMT	<---	IP	1.000			
Brixton	<---	EF	.066	.018	3.766	***
LNS	<---	WM	1.000			
Digits	<---	WM	.597	.158	3.786	***

Standardized Regression Weights:

			Estimate
AVLT7	<---	DR	.943
SMTrecall	<---	DR	.565
HAYTime	<---	EF	.755
ErrA	<---	EF	.717
ErrB	<---	EF	.782
SDMT	<---	IP	.883
Brixton	<---	EF	.452
LNS	<---	WM	.850
Digits	<---	WM	.591

Covariances:

			Estimate	S.E.	C.R.	P
DR	<-->	EF	68.120	18.288	3.725	***
DR	<-->	IP	18.436	4.921	3.746	***
EF	<-->	IP	256.490	64.940	3.950	***
IP	<-->	WM	12.978	3.357	3.866	***
DR	<-->	WM	3.712	.993	3.738	***
EF	<-->	WM	26.096	11.290	2.312	.021

Correlations:

			Estimate
DR	<-->	EF	.534
DR	<-->	IP	.546
EF	<-->	IP	.621
IP	<-->	WM	.595
DR	<-->	WM	.552
EF	<-->	WM	.317

Variances:

	Estimate	S.E.	C.R.	P
DR	10.421	2.801	3.720	***
EF	1558.838	399.950	3.898	***
IP	109.490	21.053	5.201	***
WM	4.341	1.296	3.350	***
e1	30.781			
e2	2.881	.559	5.150	***
e3	1.295	2.159	.600	.549
e4	9.131	1.668	5.475	***
e6	1174.297	250.143	4.695	***
e7	.860	.174	4.939	***
e8	2.174	.554	3.921	***
e5	26.817	4.410	6.081	***
e9	1.665	.993	1.678	.093

Squared Multiple Correlations:

	Estimate
LNS	.723
Brixton	.204
Digits	.349
SDMT	.781
ErrB	.611
ErrA	.514
HAYTime	.570
SMTrecall	.319
AVLT7	.889

Sample Moments

Sample Covariances

	LNS	Brixton	Digits	SDMT	ErrB	ErrA	HAYTime	SMTrecall	AVLT7
LNS	6.545								
Brixton	3.141	38.735							
Digits	2.550	1.372	4.540						
SDMT	14.555	29.043	7.570	152.365					
ErrB	.751	3.402	.990	8.959	6.054				
ErrA	.870	2.092	.460	6.160	1.759	1.905			
HAYTime	13.867	99.076	15.780	254.354	77.538	35.509	2838.428		
SMTrecall	2.690	6.441	1.979	15.103	.917	1.889	47.646	16.462	
AVLT7	4.617	8.723	2.170	22.784	2.068	1.949	55.838	8.511	13.608

Sample Correlations

	LNS	Brixton	Digits	SDMT	ErrB	ErrA	HAYTime	SMTrecall	AVLT7
LNS	1.000								
Brixton	.197	1.000							
Digits	.468	.103	1.000						
SDMT	.461	.378	.288	1.000					
ErrB	.119	.222	.189	.295	1.000				
ErrA	.246	.244	.156	.362	.518	1.000			
HAYTime	.102	.299	.139	.387	.591	.483	1.000		
SMTrecall	.259	.255	.229	.302	.092	.337	.220	1.000	
AVLT7	.489	.380	.276	.500	.228	.383	.284	.569	1.000

Matrices

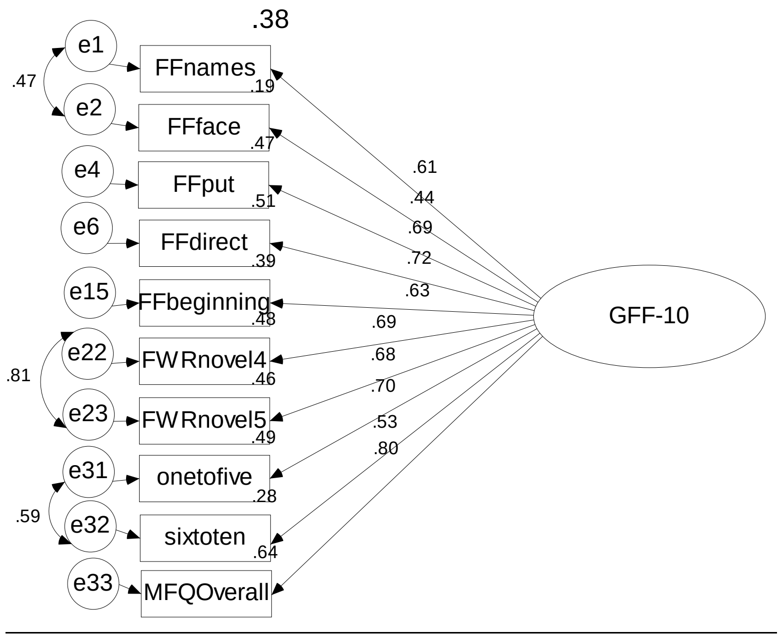
Residual Covariances

	LNS	Brixton	Digits	SDMT	ErrB	ErrA	HAYTime	SMTrecall	AVLT7
LNS	.539								
Brixton	1.406	5.029							
Digits	-.039	.338	.114						
SDMT	1.577	11.992	-.172	12.095					
ErrB	-.471	-1.449	.261	-3.049	.464				
ErrA	.240	-.409	.084	-.032	-.003	.136			
HAYTime	-12.230	-4.553	.213	-2.137	4.560	-2.123	105.293		
SMTrecall	.310	3.538	.559	3.282	-1.128	.835	3.969	3.047	
AVLT7	.904	4.194	-.045	4.347	-1.121	.304	-12.283	1.829	1.891

Standardized Residual Covariances

	LNS	Brixton	Digits	SDMT	ErrB	ErrA	HAYTime	SMTrecall	AVLT7
LNS	.631								
Brixton	.976	1.050							
Digits	-.068	.274	.182						
SDMT	.493	1.684	-.065	.607					
ErrB	-.791	-.990	.517	-.996	.583				
ErrA	.719	-.502	.297	-.019	-.008	.542			
HAYTime	-.931	-.141	.019	-.032	.316	-.267	.271		
SMTrecall	.332	1.640	.710	.726	-1.261	1.667	.201	1.598	
AVLT7	.981	2.048	-.059	.971	-1.282	.625	-.638	1.281	1.135

APPENDIX J - 1 FACTOR GENERAL FREQUENCY OF FORGETTING MODEL



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

MFQOverall	Overall how is your Memory
sixtoten	Memory compared to 6-10 years ago
onetofive	Memory compared to 1-5 years ago
FWRnovel5	Forgetting while reading a book/novel - opening chapters
FWRnovel4	Forgetting while reading a book/novel - chapter before
FFbeginning	Beginning to do something and forgetting what
FFdirect	Forgetting directions
FFput	Forgetting where you put things
FFfaces	Forgetting faces
FFnames	Forgetting names

Unobserved, exogenous variables

GFF-10	General Frequency of Forgetting

Variable counts (including residuals)

Number of variables in your model:	21
Number of observed variables:	10
Number of unobserved variables:	11
Number of exogenous variables:	11
Number of endogenous variables:	10

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	11	0	0	0	0	11
Labeled	0	0	0	0	0	0
Unlabeled	9	3	11	0	0	23
Total	20	3	11	0	0	34

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	23
Degrees of freedom (55 - 23):	32

Result

Minimum was achieved
Chi-square = 50.744
Degrees of freedom = 32
Probability level = .019

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
MFQ model 1	23	50.744	32	.019	1.586
Saturated model	55	.000	0		
Independence model	10	138.949	45	.000	3.088
Zero model	0	495.000	55	.000	9.000

GFI

Model	GFI
MFQ model 1	.897
Saturated model	1.000
Independence model	.719
Zero model	.000

Baseline Comparisons

Model	CFI
MFQ model 1	.800
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
MFQ model 1	.077	.032	.115	.136
Independence model	.145	.118	.173	.000

Regression Weights:

			Estimate	S.E.	C.R.	P
onetofive	<---	GFF-10	1.019	.190	5.357	***
FWRnovel5	<---	GFF-10	1.293	.203	6.378	***
FWRnovel4	<---	GFF-10	1.282	.201	6.372	***
FFbeginning	<---	GFF-10	1.164	.192	6.053	***
FFdirect	<---	GFF-10	1.200	.194	6.189	***
FFput	<---	GFF-10	1.008	.169	5.947	***
FFfaces	<---	GFF-10	.653	.156	4.177	***
FFnames	<---	GFF-10	.864	.162	5.346	***
MFQOverall	<---	GFF-10	1.000			
sixtoten	<---	GFF-10	.871	.221	3.949	***

Standardized Regression Weights:

			Estimate
onetofive	<---	GFF-10	.698
FWRnovel5	<---	GFF-10	.678
FWRnovel4	<---	GFF-10	.691
FFbeginning	<---	GFF-10	.628
FFdirect	<---	GFF-10	.716
FFput	<---	GFF-10	.686
FFfaces	<---	GFF-10	.438
FFnames	<---	GFF-10	.612
MFQOverall	<---	GFF-10	.798
sixtoten	<---	GFF-10	.527

Residual Covariances:

			Estimate	S.E.	C.R.	P
e2	<-->	e1	.556	.156	3.566	***
e23	<-->	e22	1.218	.248	4.901	***
e32	<-->	e31	.692	.220	3.145	.002

Residual Correlations:

			Estimate
e2	<-->	e1	.466
e23	<-->	e22	.811
e32	<-->	e31	.591

Variances:

	Estimate	S.E.	C.R.	P
GFF-10	.798	.185	4.315	***
e33	.455	.097	4.706	***
e32	1.570	.322	4.881	***
e31	.873	.199	4.378	***
e23	1.571	.274	5.732	***
e22	1.436	.264	5.450	***
e15	1.658	.281	5.901	***
e6	1.092	.217	5.025	***
e4	.910	.179	5.076	***
e2	1.437	.223	6.438	***
e1	.992	.183	5.419	***

Squared Multiple Correlations

	Estimate
FFnames	.375
FFfaces	.191
FFput	.471
FFdirect	.513
FFbeginning	.395
FWRnovel4	.477
FWRnovel5	.459
onetofive	.487
sixtoten	.278
MFQOverall	.637

Sample Covariances

	FFnames	FFfaces	FFput	FFdirect	FFbeginning	FWRnovel4	FWRnovel5	onetofive	sixtoten	MFQOverall
FFnames	1.932									
FFfaces	.977	1.941								
FFput	.728	.481	2.104							
FFdirect	.511	.461	.877	2.577						
FFbeginning	.962	.503	1.122	.897	3.030					
FWRnovel4	.531	.804	1.056	1.289	.971	3.075				
FWRnovel5	.684	.780	1.192	1.149	1.125	2.767	3.103			
onetofive	.860	.394	.386	1.264	.612	1.051	.975	2.278		
sixtoten	.750	.271	-.006	1.271	.498	.700	.613	2.122	3.228	
MFQOverall	.702	.678	.794	.868	1.049	1.043	1.044	.643	.592	1.330

Sample Correlations

	FFnames	FFfaces	FFput	FFdirect	FFbeginning	FWRnovel4	FWRnovel5	onetofive	sixtoten	MFQOverall
FFnames	1.000									
FFfaces	.505	1.000								
FFput	.361	.238	1.000							
FFdirect	.229	.206	.376	1.000						
FFbeginning	.398	.207	.444	.321	1.000					
FWRnovel4	.218	.329	.415	.458	.318	1.000				
FWRnovel5	.279	.318	.467	.406	.367	.896	1.000			
onetofive	.410	.188	.176	.522	.233	.397	.367	1.000		
sixtoten	.301	.108	-.002	.441	.159	.222	.194	.783	1.000	
MFQOverall	.438	.422	.475	.469	.523	.516	.514	.370	.286	1.000

Matrices

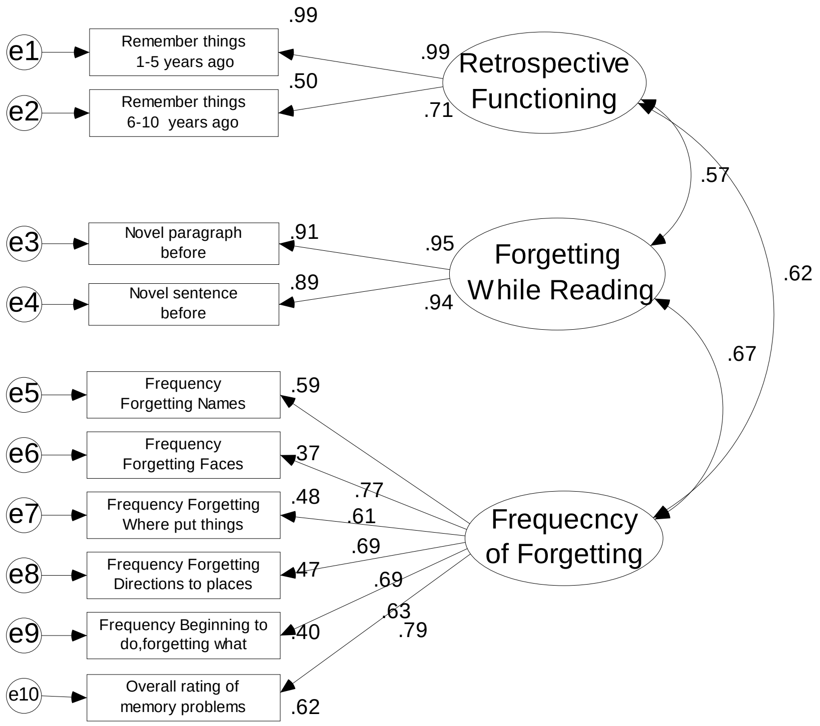
Residual Covariances

	FFnames	FFfaces	FFput	FFdirect	FFbeginning	FWRnovel4	FWRnovel5	onetofive	sixtoten	MFQOverall
FFnames	.343									
FFfaces	-.029	.163								
FFput	.033	-.044	.383							
FFdirect	-.317	-.164	-.088	.336						
FFbeginning	.160	-.104	.186	-.217	.291					
FWRnovel4	-.353	.136	.025	.061	-.220	.328				
FWRnovel5	-.208	.106	.152	-.088	-.075	.226	.198			
onetofive	.157	-.136	-.433	.288	-.334	.009	-.076	.577		
sixtoten	.150	-.183	-.706	.438	-.311	-.191	-.286	.722	1.052	
MFQOverall	.013	.157	-.010	-.089	.120	.020	.012	-.169	-.103	.077

Standardized Residual Covariances

	FFnames	FFfaces	FFput	FFdirect	FFbeginning	FWRnovel4	FWRnovel5	onetofive	sixtoten	MFQOverall
FFnames	1.522									
FFfaces	-.147	.647								
FFput	.182	-.238	1.566							
FFdirect	-1.529	-.782	-.401	1.056						
FFbeginning	.711	-.450	.783	-.796	.747					
FWRnovel4	-1.548	.586	.104	.220	-.731	.839				
FWRnovel5	-.889	.447	.615	-.310	-.245	.593	.480			
onetofive	.876	-.746	-2.274	1.316	-1.410	.038	-.307	2.387		
sixtoten	.764	-.904	-3.414	1.844	-1.203	-.729	-1.065	3.021	3.401	
MFQOverall	.082	.990	-.058	-.461	.579	.096	.056	-1.009	-.573	.435

APPENDIX K - 3 FACTOR GENERAL FREQUENCY OF FORGETTING MODEL



The model is recursive.

Sample size = 100

Observed, endogenous variables

MFQOverall	Overall how is your Memory?
sixtoten	Remembering things from 6-10 years ago
onetofive	Remembering things from 1-5 years ago
FWRnovel5	Forgetting while reading a book/novel - paragraph before
FWRnovel4	Forgetting while reading a book/novel - sentence before
FFbeginning	Beginning to do something and forgetting what
FFdirect	Forgetting directions
FFput	Forgetting where you put things
FFfaces	Forgetting faces
FFnames	Forgetting names

Unobserved, exogenous variables

RF	Retrospective Functioning
FWR	Forgetting While Reading
FF	Frequency of Forgetting

Variable counts (including residuals)

Number of variables in your model:	23
Number of observed variables:	10
Number of unobserved variables:	13
Number of exogenous variables:	13
Number of endogenous variables:	10

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	13	0	0	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	7	3	13	0	0	23
Total	20	3	13	0	0	36

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	23
Degrees of freedom (55 - 23):	32

Result

Minimum was achieved
Chi-square = 59.906
Degrees of freedom = 32
Probability level = .002

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	23	59.906	32	.002	1.872
Saturated model	55	.000	0		
Independence model	10	138.949	45	.000	3.088
Zero model	0	495.000	55	.000	9.000

GFI

Model	GFI
Default model	.879
Saturated model	1.000
Independence model	.719
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.703
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.094	.056	.130	.032
Independence model	.145	.118	.173	.000

Regression Weights:

			Estimate	S.E.	C.R.	P
FWRnovel5	<---	FWR	1.000			
FWRnovel4	<---	FWR	.976	.074	13.117	***
FFbeginning	<---	FF	1.183	.193	6.136	***
FFdirect	<---	FF	1.089	.187	5.827	***
FFput	<---	FF	1.040	.171	6.079	***
FFfaces	<---	FF	.850	.163	5.205	***
sixtoten	<---	RF	1.000			
onetofive	<---	RF	1.266	.227	5.577	***
FFnames	<---	FF	1.042	.173	6.009	***
MFQOverall	<---	FF	1.000			

Standardized Regression Weights:

			Estimate
FWRnovel5	<---	FWR	.944
FWRnovel4	<---	FWR	.952
FFbeginning	<---	FF	.633
FFdirect	<---	FF	.686
FFput	<---	FF	.692
FFfaces	<---	FF	.607
sixtoten	<---	RF	.710
onetofive	<---	RF	.995
FFnames	<---	FF	.765
MFQOverall	<---	FF	.788

Covariances:

			Estimate	S.E.	C.R.	P
FWR	<-->	FF	.962	.214	4.493	***
FWR	<-->	RF	.924	.282	3.281	.001
FF	<-->	RF	.558	.181	3.090	.002

Correlations:

			Estimate
FWR	<-->	FF	.675
FWR	<-->	RF	.569
FF	<-->	RF	.621

Variances:

	Estimate	S.E.	C.R.	P
FWR	2.580	.451	5.717	***
FF	.788	.183	4.303	***
RF	1.023	.357	2.870	.004
e2	.316	.158	2.002	.045
e1	.254	.150	1.687	.092
e10	.479	.095	5.027	***
e7	1.649	.281	5.875	***
e6	1.051	.217	4.839	***
e5	.925	.180	5.142	***
e4	.976	.186	5.249	***
e3	.605	.156	3.875	***
e9	1.008	.235	4.298	***
e8	.018	.241	.075	.940

Squared Multiple Correlations:

	Estimate
onetofive	.989
sixtoten	.504
FFnames	.586
FFfaces	.368
FFput	.479
FFdirect	.471
FFbeginning	.401
MFQOverall	.622
FWRnovel4	.906
FWRnovel5	.891

Sample Moments

Sample Covariances

	onetofive	sixtoten	FFnames	FFfaces	FFput	FFdirect	FFbeginning	MFQOverall	FWRnovel4	FWRnovel5
onetofive	2.278									
sixtoten	2.122	3.228								
FFnames	.860	.750	1.932							
FFfaces	.394	.271	.977	1.941						
FFput	.386	-.006	.728	.481	2.104					
FFdirect	1.264	1.271	.511	.461	.877	2.577				
FFbeginning	.612	.498	.962	.503	1.122	.897	3.030			
MFQOverall	.643	.592	.702	.678	.794	.868	1.049	1.330		
FWRnovel4	1.051	.700	.531	.804	1.056	1.289	.971	1.043	3.075	
FWRnovel5	.975	.613	.684	.780	1.192	1.149	1.125	1.044	2.767	3.103

Sample Correlations

	onetofive	sixtoten	FFnames	FFfaces	FFput	FFdirect	FFbeginning	MFQOverall	FWRnovel4	FWRnovel5
onetofive	1.000									
sixtoten	.783	1.000								
FFnames	.410	.301	1.000							
FFfaces	.188	.108	.505	1.000						
FFput	.176	-.002	.361	.238	1.000					
FFdirect	.522	.441	.229	.206	.376	1.000				
FFbeginning	.233	.159	.398	.207	.444	.321	1.000			
MFQOverall	.370	.286	.438	.422	.475	.469	.523	1.000		
FWRnovel4	.397	.222	.218	.329	.415	.458	.318	.516	1.000	
FWRnovel5	.367	.194	.279	.318	.467	.406	.367	.514	.896	1.000

Matrices

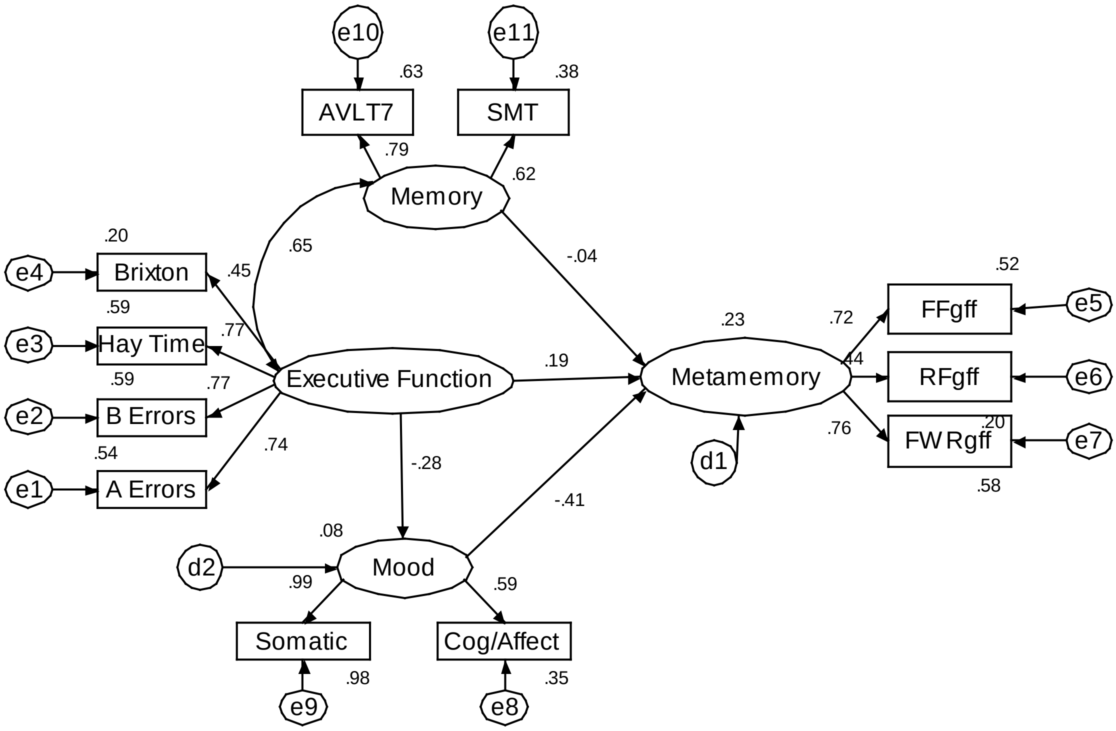
Residual Covariances

	onetofive	sixtoten	FFnames	FFfaces	FFput	FFdirect	FFbeginning	MFQOverall	FWRnovel4	FWRnovel5
onetofive	.619									
sixtoten	.827	1.196								
FFnames	.123	.169	.471							
FFfaces	-.206	-.204	.280	.396						
FFput	-.349	-.586	-.126	-.215	.327					
FFdirect	.494	.664	-.384	-.268	-.015	.592				
FFbeginning	-.224	-.162	-.009	-.289	.153	-.118	.279			
MFQOverall	-.063	.034	-.119	.009	-.025	.010	.117	.063		
FWRnovel4	-.091	-.201	-.447	.006	.080	.267	-.139	.105	.365	
FWRnovel5	-.195	-.311	-.319	-.037	.192	.102	-.012	.082	.249	.207

Standardized Residual Covariances

	onetofive	sixtoten	FFnames	FFfaces	FFput	FFdirect	FFbeginning	MFQOverall	FWRnovel4	FWRnovel5
onetofive	2.624									
sixtoten	3.660	4.142								
FFnames	.712	.925	2.269							
FFfaces	-1.199	-1.104	1.679	1.805						
FFput	-1.859	-2.934	-.688	-1.189	1.296					
FFdirect	2.495	3.148	-1.984	-1.407	-.073	2.097				
FFbeginning	-.970	-.656	-.041	-1.302	.631	-.460	.713			
MFQOverall	-.390	.199	-.744	.056	-.144	.055	.559	.349		
FWRnovel4	-.375	-.797	-2.008	.028	.332	1.047	-.470	.502	.947	
FWRnovel5	-.782	-1.193	-1.387	-.163	.771	.388	-.040	.381	.657	.503

Appendix L 10 ITEM GENERAL FREQUENCY OF FORGETTING



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
FWR	Forgetting While Reading
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FF	Frequency of Forgetting factor score
RF	Retrospective Function factor score
FWR	Forgetting while Reading factor score

Unobserved, endogenous variables

Mood
Metamemory

Unobserved, exogenous variables

Executive Function
Memory

Variable counts (including residuals)

Number of variables in your model:	28
Number of observed variables:	11
Number of unobserved variables:	17
Number of exogenous variables:	15
Number of endogenous variables:	13

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	17	0	1	0	0	18
Labeled	0	0	0	0	0	0
Unlabeled	11	1	14	0	0	26
Total	28	1	15	0	0	44

Computation of degrees of freedom

Number of distinct sample moments:	66
Number of distinct parameters to be estimated:	26
Degrees of freedom (66 - 26):	40

Result

Minimum was achieved
Chi-square = 49.035
Degrees of freedom = 40
Probability level = .155

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	26	49.035	40	.155	1.226
Saturated model	66	.000	0		
Independence model	11	128.082	55	.000	2.329
Zero model	0	544.500	66	.000	8.250

GFI

Model	GFI
Default model	.910
Saturated model	1.000
Independence model	.765
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.876
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.048	.000	.088	.503
Independence model	.116	.090	.142	.000

Regression Weights

			Estimate	S.E.	C.R.	P
Mood	<---	Executive Function	-.004	.002	-1.703	.089
MM	<---	Mood	-.452	.170	-2.661	.008
MM	<---	Memory	-.009	.067	-.126	.899
MM	<---	Executive Function	.003	.004	.776	.438
SMTrecall	<---	Executive Function	.025	.006	4.451	***
AVLT7	<---	Executive Function	.046	.009	5.215	***
FWR	<---	Executive Function	1.000			
RF	<---	Executive Function	.063	.021	3.036	.002
FF	<---	Mood	1.000			
Somatic	<---	Mood	1.493	.217	6.876	***
Cog/Affect	<---	MM	1.000			
ErrBRIX	<---	MM	.607	.195	3.112	.002
TimeHAY	<---	MM	1.166	.288	4.045	***
ErrB	<---	Memory	1.000			
ErrA	<---	Memory	.763	.305	2.504	.012

Standardized Regression Weights

			Estimate
Mood	<---	Executive Function	-.281
MM	<---	Mood	-.408
MM	<---	Memory	-.037
MM	<---	Executive Function	.191
SMTrecall	<---	Executive Function	.735
AVLT7	<---	Executive Function	.769
FWR	<---	Executive Function	.769
RF	<---	Executive Function	.445
FF	<---	Mood	.595
Somatic	<---	Mood	.992
Cog/Affect	<---	MM	.718
ErrBRIX	<---	MM	.442
TimeHAY	<---	MM	.764
ErrB	<---	Memory	.794
ErrA	<---	Memory	.618

Covariances

	Estimate	S.E.	C.R.	P
Executive Function <--> Memory	65.507	21.814	3.003	.003

Correlations

	Estimate
Executive Function <--> Memory	.647

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1563.621	459.660	3.402	***
Memory	6.565	2.621	2.505	.012
d2	.261	.087	3.020	.003
d1	.268	.099	2.701	.007
e9	.010			
e1	.801	.174	4.612	***
e2	2.333	.555	4.200	***
e3	1079.951	264.782	4.079	***
e4	25.018	4.298	5.821	***
e8	.518	.075	6.897	***
e5	.328	.088	3.722	***
e6	.532	.092	5.797	***
e7	.338	.116	2.915	.004
e1o	3.856	2.061	1.871	.061
e11	6.179	1.628	3.796	***

Squared Multiple Correlations

	Estimate
Mood	.079
MM	.233
SMTrecall	.382
AVLT7	.630
FWR	.584
RF	.195
FF	.516
Somatic	.984
Cog/Affect	.354
ErrBRIX	.198
TimeHAY	.591
ErrB	.591
ErrA	.541

Sample Moments

Sample Covariances

	SMTrecall	AVLT7	FWR	RF	FF	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	16.462										
AVLT7	8.511	13.608									
FWR	.413	-.015	.935								
RF	1.358	.456	.289	.869							
FF	.572	.089	.501	.332	.806						
Somatic	.459	.230	-.226	-.083	-.185	.774					
Cog/Affect	.414	.341	-.158	-.074	-.156	.516	.881				
ErrBRIX	6.441	8.723	.394	1.132	1.154	-.775	-.459	38.735			
TimeHAY	40.196	52.428	14.229	4.149	10.639	-6.606	-6.363	90.076	2838.428		
ErrB	.917	2.068	.234	.058	.105	-.533	-.372	3.402	74.968	6.054	
ErrA	1.889	1.949	.079	.128	.125	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

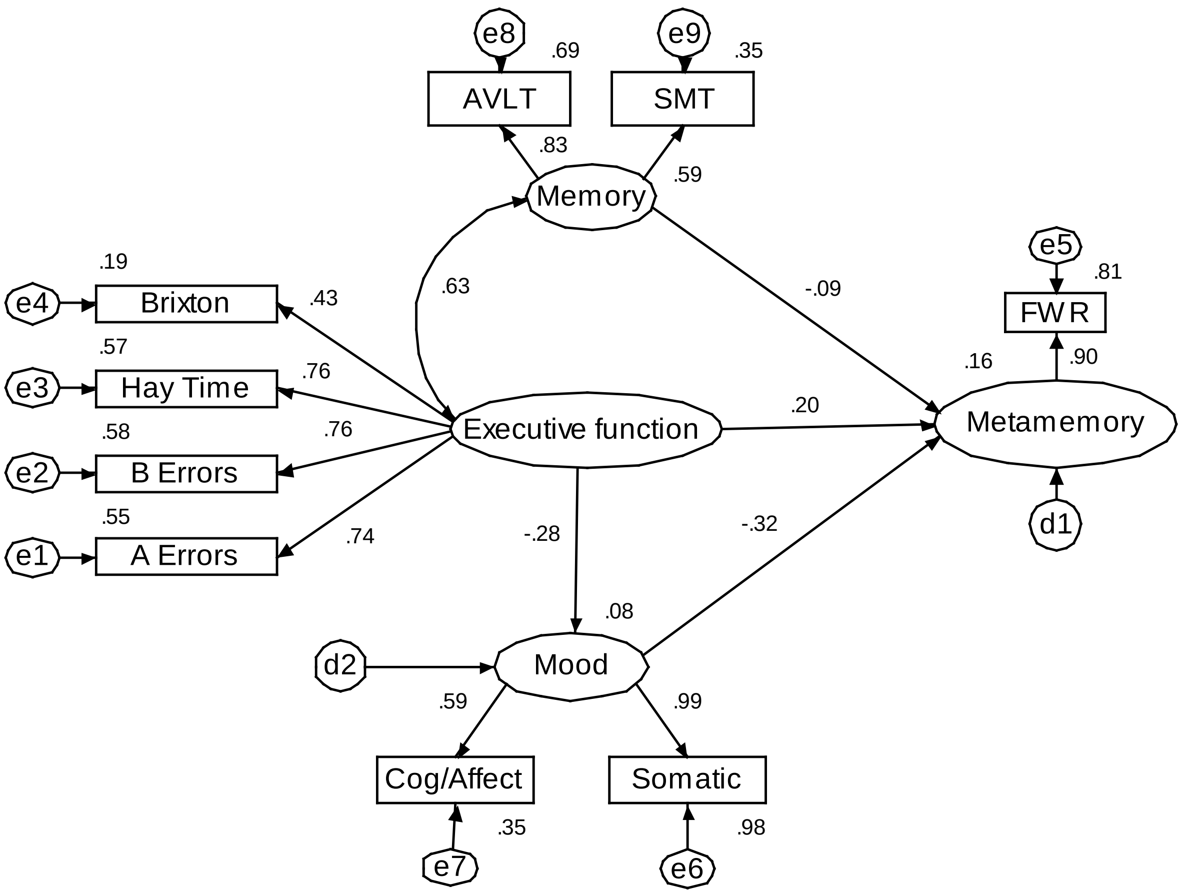
	SMTrecall	AVLT7	FWR	RF	FF	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	1.000										
AVLT7	.569	1.000									
FWR	.105	-.004	1.000								
RF	.359	.133	.321	1.000							
FF	.157	.027	.578	.397	1.000						
Somatic	.129	.071	-.266	-.102	-.234	1.000					
Cog/Affect	.109	.098	-.174	-.085	-.185	.625	1.000				
ErrBRIX	.255	.380	.065	.195	.206	-.142	-.079	1.000			
TimeHAY	.186	.267	.276	.084	.222	-.141	-.127	.272	1.000		
ErrB	.092	.228	.098	.025	.048	-.246	-.161	.222	.572	1.000	
ErrA	.337	.383	.059	.099	.101	-.226	-.116	.244	.445	.518	1.000

Residual Covariances

	SMTrecall	AVLT7	FWR	RF	FF	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	6.465										
AVLT7	3.504	3.186									
FWR	.196	-.299	.121								
RF	1.245	.308	.041	.208							
FF	.386	-.155	.094	.120	.128						
Somatic	.742	.601	.023	.046	.029	.132					
Cog/Affect	.603	.589	.009	.013	-.012	.093	.079				
ErrBRIX	3.301	4.605	-.090	.880	.739	-.219	-.087	7.539			
TimeHAY	-9.759	-13.080	6.535	.142	4.041	2.233	-.441	-8.209	194.855		
ErrB	-1.402	-.973	-.123	-.128	-.201	-.123	-.098	-1.160	2.381	.352	
ErrA	.662	.339	-.110	.029	-.037	-.057	-.004	-.322	-5.724	-.024	.161

Standardized Residual Covariances

	SMTrecall	AVLT7	FWR	RF	FF	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	4.550										
AVLT7	3.067	2.151									
FWR	.681	-1.017	1.043								
RF	4.818	1.165	.532	2.217							
FF	1.470	-.576	1.098	1.694	1.329						
Somatic	2.895	2.287	.297	.695	.413	1.442					
Cog/Affect	2.116	2.020	.108	.174	-.165	1.109	.697				
ErrBRIX	1.831	2.477	-.176	1.925	1.593	-.484	-.172	1.700			
TimeHAY	-.571	-.729	1.383	.034	.938	.527	-.095	-.269	.519		
ErrB	-1.767	-1.168	-.560	-.654	-1.006	-.624	-.450	-.819	.166	.435	
ErrA	1.514	.741	-.906	.271	-.333	-.526	-.036	-.413	-.730	-.067	.647



Title: Forgetting While Reading

The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
FWR	Forgetting While Reading
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall

Unobserved, endogenous variables
Mood
Metamemory

Unobserved, exogenous variables
Executive function
Memory

Variable counts (including residuals)

Number of variables in your model:	24
Number of observed variables:	9
Number of unobserved variables:	15
Number of exogenous variables:	13
Number of endogenous variables:	11

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	15	0	2	0	0	17
Labeled	0	0	0	0	0	0
Unlabeled	9	1	11	0	0	21
Total	24	1	13	0	0	38

Computation of degrees of freedom

Number of distinct sample moments:	45
Number of distinct parameters to be estimated:	21
Degrees of freedom (45 - 21):	24

Result (Default model)

Minimum was achieved
Chi-square = 36.125
Degrees of freedom = 24
Probability level = .053

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	21	36.125	24	.053	1.505
Saturated model	45	.000	0		
Independence model	9	101.000	36	.000	2.806
Zero model	0	445.500	45	.000	9.900

GFI

Model	GFI
Default model	.919
Saturated model	1.000
Independence model	.773
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.813
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.071	.000	.117	.219
Independence model	.135	.104	.167	.000

Regression Weights

			Estimate	S.E.	C.R.	P
Mood	<---	Executive function	-.227	.125	-1.821	.069
MM	<---	Executive function	.167	.190	.880	.379
MM	<---	Mood	-.327	.131	-2.492	.013
MM	<---	Memory	-.027	.072	-.369	.712
ErrA	<---	Executive function	1.000			
ErrB	<---	Executive function	1.854	.393	4.720	***
TimeHAY	<---	Executive function	39.512	9.321	4.239	***
ErrBRIX	<---	Executive function	2.548	.885	2.878	.004
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	.664	.098	6.800	***
FWR	<---	MM	1.000			
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.755	.296	2.549	.011

Standardized Regression Weights

			Estimate
Mood	<---	Executive function	-.282
MM	<---	Executive function	.203
MM	<---	Mood	-.320
MM	<---	Memory	-.088
ErrA	<---	Executive function	.742
ErrB	<---	Executive function	.761
TimeHAY	<---	Executive function	.757
ErrBRIX	<---	Executive function	.434
Cog/Affect	<---	Mood	.992
Somatic	<---	Mood	.590
FWR	<---	MM	.899
AVLT7	<---	Memory	.828
SMTrecall	<---	Memory	.588

Covariances

	Estimate	S.E.	C.R.	P
Executive function <--> Memory	1.676	.660	2.540	.011

Correlations

	Estimate
Executive function <--> Memory	.633

Variances

	Estimate	S.E.	C.R.	P
Executive function	.967	.308	3.139	.002
Memory	7.240	2.779	2.605	.009
d2	.580	.095	6.104	***
d1	.555	.115	4.833	***
e6	.010			
e5	.156			
e1	.789	.177	4.467	***
e2	2.423	.558	4.340	***
e3	1126.006	263.154	4.279	***
e4	27.131	4.487	6.046	***
e7	.522	.075	6.918	***
e8	3.322	2.266	1.466	.143
e9	7.799	1.809	4.311	***

Squared Multiple Correlations

	Estimate
Mood	.079
MM	.156
SMTrecall	.346
AVLT7	.685
FWR	.808
Cog/Affect	.348
Somatic	.984
ErrBRIX	.188
TimeHAY	.573
ErrB	.579
ErrA	.551

Sample Moments

Sample Covariances

	SMTrecall	AVLT7	FWR	Cog/Affect	Somatic	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	16.462								
AVLT7	8.511	13.608							
FWR	.340	-.240	.967						
Cog/Affect	.414	.341	-.151	.881					
Somatic	.459	.230	-.209	.516	.774				
ErrBRIX	6.441	8.723	.265	-.459	-.775	38.735			
TimeHAY	40.196	52.428	14.889	-6.363	-6.606	90.076	2838.428		
ErrB	.917	2.068	.319	-.372	-.533	3.402	74.968	6.054	
ErrA	1.889	1.949	.085	-.150	-.274	2.092	32.689	1.759	1.905

Sample Correlations

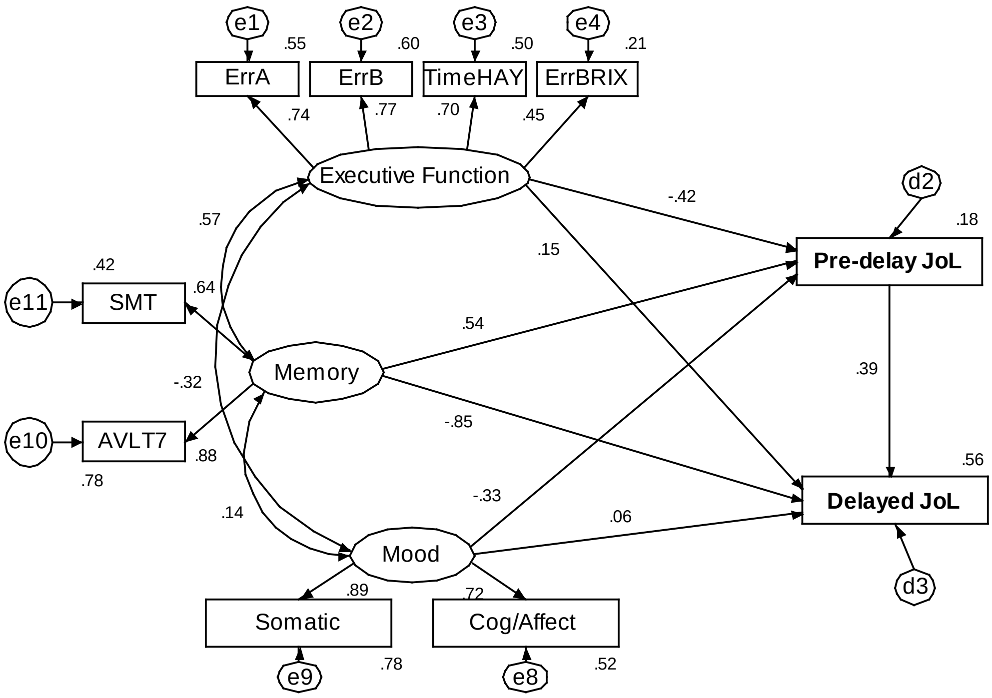
	SMTrecall	AVLT7	FWR	Cog/Affect	Somatic	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	1.000								
AVLT7	.569	1.000							
FWR	.085	-.066	1.000						
Cog/Affect	.109	.098	-.163	1.000					
Somatic	.129	.071	-.242	.625	1.000				
ErrBRIX	.255	.380	.043	-.079	-.142	1.000			
TimeHAY	.186	.267	.284	-.127	-.141	.272	1.000		
ErrB	.092	.228	.132	-.161	-.246	.222	.572	1.000	
ErrA	.337	.383	.063	-.116	-.226	.244	.445	.518	1.000

Residual Covariances

	SMTrecall	AVLT7	FWR	Cog/Affect	Somatic	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	4.533								
AVLT7	3.044	3.046							
FWR	.179	-.453	.153						
Cog/Affect	.605	.594	.004	.082					
Somatic	.747	.611	.023	.098	.134				
ErrBRIX	3.217	4.453	-.217	-.087	-.215	5.323			
TimeHAY	-9.803	-13.781	7.407	-.597	2.075	-7.309	202.397		
ErrB	-1.430	-1.039	-.032	-.102	-.126	-1.168	4.100	.305	
ErrA	.624	.273	-.104	-.004	-.055	-.372	-5.529	-.035	.149

Standardized Residual Covariances

	SMTrecall	AVLT7	FWR	Cog/Affect	Somatic	ErrBRIX	TimeHAY	ErrB	ErrA
SMTrecall	2.674								
AVLT7	2.426	2.029							
FWR	.571	-1.533	1.327						
Cog/Affect	1.946	2.025	.049	.721					
Somatic	2.675	2.314	.308	1.178	1.472				
ErrBRIX	1.583	2.300	-.413	-.167	-.459	1.121			
TimeHAY	-.529	-.764	1.571	-.128	.492	-.233	.540		
ErrB	-1.653	-1.233	-.145	-.469	-.637	-.796	.287	.374	
ErrA	1.307	.587	-.855	-.032	-.501	-.460	-.705	-.095	.595



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
Pre-delay JoL	Trail 5 JoL
Delayed JoL	Trial 7 JoL

Unobserved, exogenous variables

Executive Function
Mood
Memory

Variable counts (including residuals)

Number of variables in your model:	23
Number of observed variables:	10
Number of unobserved variables:	13
Number of exogenous variables:	13
Number of endogenous variables:	10

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	13	0	0	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	12	3	13	0	0	28
Total	25	3	13	0	0	41

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	28
Degrees of freedom (55 - 28):	27

Result

Minimum was achieved
Chi-square = 30.972
Degrees of freedom = 27
Probability level = .272

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	28	30.972	27	.272	1.147
Saturated model	55	.000	0		
Independence model	10	123.170	45	.000	2.737
Zero model	0	495.000	55	.000	9.000

GFI

Model	GFI
Default model	.937
Saturated model	1.000
Independence model	.751
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.949
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.039	.000	.091	.588
Independence model	.132	.105	.161	.000

Regression Weights

			Estimate	S.E.	C.R.	P
Pre-delay JoL	<---	Executive Function	-.026	.016	-1.613	.107
Pre-delay JoL	<---	Memory	.401	.188	2.133	.033
Pre-delay JoL	<---	Mood	-1.096	.541	-2.028	.043
ErrA	<---	Executive Function	.027	.006	4.727	***
ErrB	<---	Executive Function	.049	.008	5.922	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.071	.025	2.831	.005
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.136	.322	3.525	***
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.771	.162	4.747	***
Delayed JoL	<---	Executive Function	.010	.014	.666	.505
Delayed JoL	<---	Memory	-.678	.200	-3.384	***
Delayed JoL	<---	Mood	.218	.521	.419	.676
Delayed JoL	<---	Pre-delay JoL	.419	.120	3.508	***

Standardized Regression Weights

			Estimate
Pre-delay JoL	<---	Executive Function	-.420
Pre-delay JoL	<---	Memory	.542
Pre-delay JoL	<---	Mood	-.329
ErrA	<---	Executive Function	.740
ErrB	<---	Executive Function	.774
TimeHAY	<---	Executive Function	.705
ErrBRIX	<---	Executive Function	.453
Cog/Affect	<---	Mood	.722
Somatic	<---	Mood	.885
AVLT7	<---	Memory	.884
SMTrecall	<---	Memory	.645
Delayed JoL	<---	Executive Function	.145
Delayed JoL	<---	Memory	-.846
Delayed JoL	<---	Mood	.060
Delayed JoL	<---	Pre-delay JoL	.387

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	64.105	19.527	3.283	.001
Mood	<-->	Memory	.283	.288	.984	.325
Executive Function	<-->	Mood	-7.963	3.463	-2.300	.021

Correlations

			Estimate
Executive Function	<-->	Memory	.575
Mood	<-->	Memory	.139
Executive Function	<-->	Mood	-.322

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1356.077	412.914	3.284	.001
Mood	.451	.163	2.764	.006
Memory	9.180	2.330	3.940	***
d2	4.127	.801	5.151	***
e1	.806	.175	4.606	***
e2	2.223	.569	3.910	***
e3	1374.618	272.190	5.050	***
e4	26.160	4.371	5.985	***
e8	.415	.128	3.247	.001
e9	.161	.150	1.072	.284
e1o	2.570	1.201	2.140	.032
e11	7.666	1.405	5.456	***
d3	2.610	.690	3.781	***

Squared Multiple Correlations

	Estimate
Pre-delay JoL	.178
Delayed JoL	.557
SMTrecall	.416
AVLT7	.781
Somatic	.784
Cog/Affect	.521
ErrBRIX	.205
TimeHAY	.497
ErrB	.598
ErrA	.548

Sample Moments

Sample Covariances

	Pre-delay JoL	Delayed JoL	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	5.780									
Delayed JoL	1.526	6.870								
SMTrecall	2.494	-3.570	16.462							
AVLT7	.999	-6.115	8.511	13.608						
Somatic	-.094	-.157	.459	.230	.774					
Cog/Affect	-.286	-.353	.414	.341	.516	.881				
ErrBRIX	-1.374	-4.149	6.441	8.723	-.775	-.459	38.735			
TimeHAY	-4.381	-24.365	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	.143	-1.008	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	.071	-1.131	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	Pre-delay JoL	Delayed JoL	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	1.000									
Delayed JoL	.242	1.000								
SMTrecall	.256	-.336	1.000							
AVLT7	.113	-.632	.569	1.000						
Somatic	-.044	-.068	.129	.071	1.000					
Cog/Affect	-.127	-.144	.109	.098	.625	1.000				
ErrBRIX	-.092	-.254	.255	.380	-.142	-.079	1.000			
TimeHAY	-.034	-.174	.186	.267	-.141	-.127	.272	1.000		
ErrB	.024	-.156	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	.022	-.313	.337	.383	-.226	-.116	.244	.445	.518	1.000

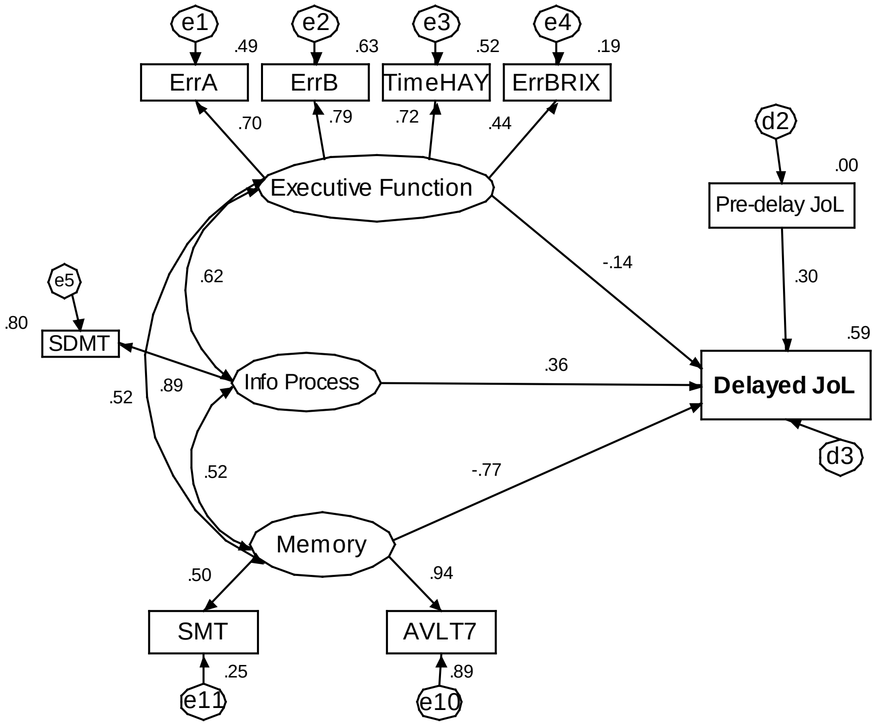
Residual Covariances

	Pre-delay JoL	Delayed JoL	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	.757									
Delayed JoL	.635	.980								
SMTrecall	1.160	.146	3.345							
AVLT7	-.732	-1.293	1.438	1.858						
Somatic	.108	.121	.212	-.091	.030					
Cog/Affect	-.108	-.109	.196	.058	.004	.015				
ErrBRIX	-1.356	-1.869	2.956	4.200	-.137	.103	5.824			
TimeHAY	-4.126	7.947	-9.200	-11.677	2.439	1.600	-5.607	107.733		
ErrB	.156	.589	-1.524	-1.100	-.086	.021	-1.326	7.961	.521	
ErrA	.078	-.264	.564	.229	-.032	.064	-.475	-3.692	-.039	.123

Standardized Residual Covariances

	Pre-delay JoL	Delayed JoL	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	1.061									
Delayed JoL	1.146	1.171								
SMTrecall	1.403	.152	1.794							
AVLT7	-.925	-1.339	1.001	1.112						
Somatic	.552	.569	.672	-.306	.288					
Cog/Affect	-.515	-.478	.577	.179	.038	.124				
ErrBRIX	-1.050	-1.318	1.396	2.071	-.273	.191	1.245			
TimeHAY	-.351	.604	-.468	-.611	.528	.323	-.177	.278		
ErrB	.294	.988	-1.711	-1.263	-.413	.094	-.922	.566	.662	
ErrA	.260	-.784	1.120	.466	-.267	.504	-.585	-.467	-.107	.487

APPENDIX O JUDGMENT OF LEARNING STRUCTURAL MODEL, INCLUDING INFORMATION PROCESSING



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
Pre-delay JoL	Trail 5 JoL
Delayed JoL	Trial 7 JoL

Unobserved, exogenous variables

Executive Function
Memory
Info Process

Variable counts (residuals included)

Number of variables in your model:	21
Number of observed variables:	9
Number of unobserved variables:	12
Number of exogenous variables:	12
Number of endogenous variables:	9

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	12	0	1	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	8	3	11	0	0	22
Total	20	3	12	0	0	35

Computation of degrees of freedom

Number of distinct sample moments: 45
Number of distinct parameters to be estimated: 22
Degrees of freedom (45 - 22): 23

Result

Minimum was achieved
Chi-square = 32.547
Degrees of freedom = 23
Probability level = .089

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	22	32.547	23	.089	1.415
Saturated model	45	.000	0		
Independence model	9	109.160	36	.000	3.032
Zero model	0	445.500	45	.000	9.900

GFI

Model	GFI
Default model	.927
Saturated model	1.000
Independence model	.755
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.870
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.065	.000	.112	.295
Independence model	.143	.113	.174	.000

Regression Weights

			Estimate	S.E.	C.R.	P
ErrA	<---	Executive Function	.024	.005	4.730	***
ErrB	<---	Executive Function	.050	.009	5.666	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.066	.025	2.673	.008
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.536	.158	3.389	***
Delayed JoL	<---	Pre-delay JoL	.354	.091	3.894	***
SDMT	<---	Info Process	1.000			
Delayed JoL	<---	Info Process	.087	.034	2.547	.011
Delayed JoL	<---	Memory	-.621	.178	-3.484	***
Delayed JoL	<---	Executive Function	-.009	.009	-1.000	.317

Standardized Regression Weights

			Estimate
ErrA	<---	Executive Function	.703
ErrB	<---	Executive Function	.791
TimeHAY	<---	Executive Function	.720
ErrBRIX	<---	Executive Function	.438
AVLT7	<---	Memory	.945
SMTrecall	<---	Memory	.502
Delayed JoL	<---	Pre-delay JoL	.302
SDMT	<---	Info Process	.892
Delayed JoL	<---	Info Process	.359
Delayed JoL	<---	Memory	-.770
Delayed JoL	<---	Executive Function	-.139

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	61.675	22.387	2.755	.006
Memory	<-->	Info Process	16.991	5.817	2.921	.003
Executive Function	<-->	Info Process	241.058	74.659	3.229	.001

Correlations

			Estimate
Executive Function	<-->	Memory	.523
Memory	<-->	Info Process	.521
Executive Function	<-->	Info Process	.616

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1415.741	405.309	3.493	***
Memory	9.816	2.486	3.949	***
Info Process	108.163	23.311	4.640	***
d2	4.622	.740	6.247	***
e5	27.703			
e1	.869	.176	4.930	***
e2	2.151	.613	3.509	***
e3	1317.625	266.547	4.943	***
e4	25.677	4.343	5.913	***
e1o	1.182	1.673	.706	.480
e11	8.369	1.448	5.781	***
d3	2.583	.773	3.341	***

Squared Multiple Correlations

	Estimate
Pre-delay JoL	.000
SDMT	.796
Delayed JoL	.595
SMTrecall	.252
AVLT7	.893
ErrBRIX	.192
TimeHAY	.518
ErrB	.626
ErrA	.494

Sample Moments

Sample Covariances

	Pre-delay JoL	SDMT	Delayed JoL	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	5.780								
SDMT	1.021	152.365							
Delayed JoL	1.526	-5.371	6.870						
SMTrecall	2.494	15.103	-3.570	16.462					
AVLT7	.999	22.784	-6.115	8.511	13.608				
ErrBRIX	-1.374	29.043	-4.149	6.441	8.723	38.735			
TimeHAY	-4.381	257.314	-24.365	40.196	52.428	90.076	2838.428		
ErrB	.143	8.959	-1.008	.917	2.068	3.402	74.968	6.054	
ErrA	.071	6.160	-1.131	1.889	1.949	2.092	32.689	1.759	1.905

Sample Correlations

	Pre-delay JoL	SDMT	Delayed JoL	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	1.000								
SDMT	.034	1.000							
Delayed JoL	.242	-.166	1.000						
SMTrecall	.256	.302	-.336	1.000					
AVLT7	.113	.500	-.632	.569	1.000				
ErrBRIX	-.092	.378	-.254	.255	.380	1.000			
TimeHAY	-.034	.391	-.174	.186	.267	.272	1.000		
ErrB	.024	.295	-.156	.092	.228	.222	.572	1.000	
ErrA	.022	.362	-.313	.337	.383	.244	.445	.518	1.000

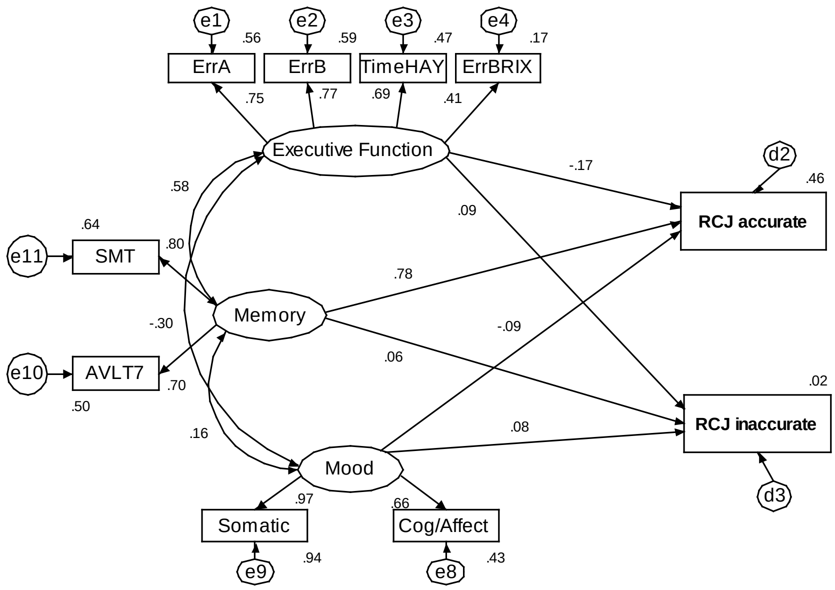
Residual Covariances

	Pre-delay JoL	SDMT	Delayed JoL	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	1.158								
SDMT	1.021	16.500							
Delayed JoL	-.112	-2.003	.495						
SMTrecall	2.494	5.996	-.789	5.272					
AVLT7	.999	5.792	-.927	3.250	2.610				
ErrBRIX	-1.374	13.233	-2.149	4.273	4.678	6.968			
TimeHAY	-4.381	16.256	6.127	7.139	-9.248	-2.779	105.061		
ErrB	.143	-3.199	.530	-.751	-1.043	-1.281	3.566	.302	
ErrA	.071	.256	-.384	1.080	.438	-.182	-1.985	.010	.187

Standardized Residual Covariances

	Pre-delay JoL	SDMT	Delayed JoL	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
Pre-delay JoL	1.762								
SDMT	.406	.854							
Delayed JoL	-.196	-.673	.546						
SMTrecall	3.451	1.490	-.883	3.315					
AVLT7	1.394	1.365	-.936	2.634	1.669				
ErrBRIX	-1.128	1.948	-1.488	2.240	2.434	1.543			
TimeHAY	-.388	.247	.450	.399	-.500	-.089	.270		
ErrB	.276	-1.044	.844	-.911	-1.215	-.891	.246	.369	
ErrA	.252	.156	-1.127	2.410	.947	-.234	-.257	.028	.766

APPENDIX P - RETROSPECTIVE CONFIDENCE STRUCTURAL MODEL CALIBRATION



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables.

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time (complex minus simple)
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
RCJ accuracy	% High RCJ correct
RCJ inaccuracy	% Low RCJ correct

Unobserved, exogenous variables

Executive Function
Mood
Memory

Variable counts (including residuals)

Number of variables in your model:	23
Number of observed variables:	10
Number of unobserved variables:	13
Number of exogenous variables:	13
Number of endogenous variables:	10

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	13	0	0	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	11	3	13	0	0	27
Total	24	3	13	0	0	40

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	27
Degrees of freedom (55 - 27):	28

Result

Minimum was achieved
Chi-square = 30.433
Degrees of freedom = 28
Probability level = .343

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	27	30.433	28	.343	1.087
Saturated model	55	.000	0		
Independence model	10	103.002	45	.000	2.289
Zero model	0	495.000	55	.000	9.000

GFI

Model	GFI
Default model	.939
Saturated model	1.000
Independence model	.792
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.958
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.030	.000	.085	.665
Independence model	.114	.085	.143	.000

Regression Weights

			Estimate	S.E.	C.R.	P
ErrA	<---	Executive Function	.028	.006	4.454	***
ErrB	<---	Executive Function	.050	.009	5.521	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.065	.028	2.347	.019
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.380	.487	2.834	.005
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	1.240	.267	4.647	***
RCJ accuracy	<---	Executive Function	-.089	.108	-.820	.412
RCJ accuracy	<---	Memory	6.155	1.925	3.198	.001
RCJ inaccuracy	<---	Executive Function	.030	.075	.401	.688
RCJ inaccuracy	<---	Memory	.295	1.025	.288	.773
RCJ accuracy	<---	Mood	-2.884	4.083	-.707	.480
RCJ inaccuracy	<---	Mood	1.562	2.855	.547	.584

Standardized Regression Weights

			Estimate
ErrA	<---	Executive Function	.745
ErrB	<---	Executive Function	.766
TimeHAY	<---	Executive Function	.686
ErrBRIX	<---	Executive Function	.408
Cog/Affect	<---	Mood	.657
Somatic	<---	Mood	.971
AVLT7	<---	Memory	.705
SMTrecall	<---	Memory	.800
RCJ accuracy	<---	Executive Function	-.165
RCJ accuracy	<---	Memory	.780
RCJ inaccuracy	<---	Executive Function	.089
RCJ inaccuracy	<---	Memory	.059
RCJ accuracy	<---	Mood	-.093
RCJ inaccuracy	<---	Mood	.080

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	49.097	21.666	2.266	.023
Mood	<-->	Memory	.232	.220	1.054	.292
Executive Function	<-->	Mood	-6.474	3.457	-1.873	.061

Correlations

			Estimate
Executive Function	<-->	Memory	.585
Mood	<-->	Memory	.159
Executive Function	<-->	Mood	-.302

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1235.489	412.629	2.994	.003
Mood	.373	.164	2.268	.023
Memory	5.707	2.169	2.631	.009
e1	.801	.179	4.464	***
e2	2.152	.566	3.801	***
e3	1387.902	270.750	5.126	***
e4	26.113	4.383	5.958	***
e8	.490	.138	3.539	***
e9	.044	.227	.192	.848
e1o	5.789	1.188	4.872	***
e11	4.942	1.430	3.456	***
d2	191.777	43.404	4.418	***
d3	138.183	20.681	6.682	***

Squared Multiple Correlations

	Estimate
RCJ inaccuracy	.021
RCJ accuracy	.461
SMTrecall	.640
AVLT7	.496
Somatic	.942
Cog/Affect	.432
ErrBRIX	.166
TimeHAY	.471
ErrB	.586
ErrA	.556

Sample Moments

Sample Covariances

	RCJ inaccuracy	RCJ accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJ inaccuracy	156.234									
RCJ accuracy	19.500	445.718								
SMTrecall	7.424	53.973	16.462							
AVLT7	6.796	36.302	8.511	13.608						
Somatic	.920	1.205	.459	.230	.774					
Cog/Affect	-.310	1.799	.414	.341	.516	.881				
ErrBRIX	2.256	37.982	6.441	8.723	-.775	-.459	38.735			
TimeHAY	50.986	93.808	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	-1.020	1.419	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	2.358	6.992	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	RCJ inaccuracy	RCJ accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJ inaccuracy	1.000									
RCJ accuracy	.074	1.000								
SMTrecall	.146	.630	1.000							
AVLT7	.147	.466	.569	1.000						
Somatic	.084	.065	.129	.071	1.000					
Cog/Affect	-.026	.091	.109	.098	.625	1.000				
ErrBRIX	.029	.289	.255	.380	-.142	-.079	1.000			
TimeHAY	.077	.083	.186	.267	-.141	-.127	.272	1.000		
ErrB	-.033	.027	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	.137	.240	.337	.383	-.226	-.116	.244	.445	.518	1.000

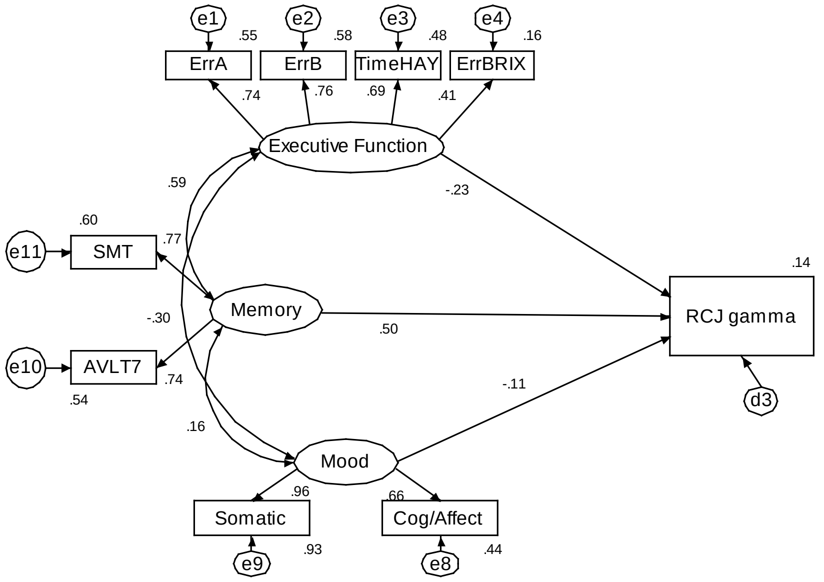
Residual Covariances

	RCJ inaccuracy	RCJ accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJ inaccuracy	15.058									
RCJ accuracy	2.838	90.105								
SMTrecall	3.062	16.657	2.748							
AVLT7	3.278	6.204	1.436	2.112						
Somatic	.289	-.071	.063	-.089	.020					
Cog/Affect	-.767	.874	.127	.109	.002	.019				
ErrBRIX	-.434	24.261	2.487	5.533	-.194	-.038	7.407			
TimeHAY	9.576	-117.389	-20.676	3.330	2.330	.112	9.810	215.037		
ErrB	-3.079	-9.079	-2.109	-.373	-.089	-.051	-.588	13.553	.849	
ErrA	1.179	.980	.157	.551	-.020	.034	-.193	-2.483	.010	.103

Standardized Residual Covariances

	RCJ inaccuracy	RCJ accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJ inaccuracy	.750									
RCJ accuracy	.126	1.783								
SMTrecall	.689	2.093	1.410							
AVLT7	.807	.874	.991	1.293						
Somatic	.279	-.043	.194	-.300	.187					
Cog/Affect	-.691	.496	.366	.344	.019	.153				
ErrBRIX	-.065	2.268	1.172	2.861	-.395	-.073	1.663			
TimeHAY	.156	-1.181	-1.033	.184	.511	.023	.328	.577		
ErrB	-1.127	-2.040	-2.339	-.457	-.436	-.235	-.437	1.021	1.148	
ErrA	.734	.375	.296	1.151	-.166	.272	-.244	-.320	.030	.402

APPENDIX Q - RETROSPECTIVE CONFIDENCE STRUCTURAL MODEL RESOLUTION



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Facto
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
RCJgamma	Retrospective Confidence Judgment gamma correlation

Unobserved, exogenous variables

Executive Function
Mood
Memory

Variable counts (including residuals)

Number of variables in your model:	21
Number of observed variables:	9
Number of unobserved variables:	12
Number of exogenous variables:	12
Number of endogenous variables:	9

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	12	0	0	0	0	12
Labeled	0	0	0	0	0	0
Unlabeled	8	3	12	0	0	23
Total	20	3	12	0	0	35

Computation of degrees of freedom

Number of distinct sample moments:	45
Number of distinct parameters to be estimated:	23
Degrees of freedom (45 - 23):	22

Result

Minimum was achieved
Chi-square = 27.975
Degrees of freedom = 22
Probability level = .177

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	23	27.975	22	.177	1.272
Saturated model	45	.000	0		
Independence model	9	98.641	36	.000	2.740
Zero model	0	445.500	45	.000	9.900

GFI

Model	GFI
Default model	.937
Saturated model	1.000
Independence model	.779
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.905
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.052	.000	.104	.436
Independence model	.133	.102	.164	.000

Regression Weights

			Estimate	S.E.	C.R.	P	Label
ErrA	<---	Executive Function	.028	.006	4.591	***	par_1
ErrB	<---	Executive Function	.050	.009	5.720	***	par_2
TimeHAY	<---	Executive Function	1.000				
ErrBRIX	<---	Executive Function	.065	.026	2.547	.011	par_3
Cog/Affect	<---	Mood	1.000				
Somatic	<---	Mood	1.363	.496	2.748	.006	par_4
AVLT7	<---	Memory	1.000				
SMTrecall	<---	Memory	1.147	.301	3.803	***	par_5
RCJgamma	<---	Executive Function	-.001	.001	-.962	.336	par_7
RCJgamma	<---	Memory	.037	.020	1.823	.068	par_8
RCJgamma	<---	Mood	-.034	.047	-.722	.471	par_9

Standardized Regression Weights

			Estimate
ErrA	<---	Executive Function	.741
ErrB	<---	Executive Function	.764
TimeHAY	<---	Executive Function	.691
ErrBRIX	<---	Executive Function	.405
Cog/Affect	<---	Mood	.663
Somatic	<---	Mood	.965
AVLT7	<---	Memory	.737
SMTrecall	<---	Memory	.773
RCJgamma	<---	Executive Function	-.233
RCJgamma	<---	Memory	.499
RCJgamma	<---	Mood	-.114

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	52.698	22.478	2.344	.019
Mood	<-->	Memory	.243	.233	1.046	.296
Executive Function	<-->	Mood	-6.590	3.511	-1.877	.061

Correlations

			Estimate
Executive Function	<-->	Memory	.594
Mood	<-->	Memory	.160
Executive Function	<-->	Mood	-.301

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1273.485	413.033	3.083	.002
Mood	.377	.171	2.208	.027
Memory	6.172	2.428	2.542	.011
e1	.822	.180	4.567	***
e2	2.231	.571	3.909	***
e3	1390.342	273.267	5.088	***
e4	27.503	4.475	6.146	***
e8	.482	.141	3.413	***
e9	.052	.229	.226	.821
e1o	5.184	1.414	3.666	***
e11	5.483	1.759	3.117	.002
d3	.029	.005	5.630	***

Squared Multiple Correlations:

	Estimate
RCJgamma	.144
SMTrecall	.597
AVLT7	.544
Somatic	.931
Cog/Affect	.439
ErrBRIX	.164
TimeHAY	.478
ErrB	.583
ErrA	.549

Sample Moments

Sample Covariances

	RCJgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJgamma	.039								
SMTrecall	.272	16.462							
AVLT7	.100	8.511	13.608						
Somatic	.005	.459	.230	.774					
Cog/Affect	.023	.414	.341	.516	.881				
ErrBRIX	.058	6.441	8.723	-.775	-.459	38.735			
TimeHAY	-.148	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	.008	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	.009	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

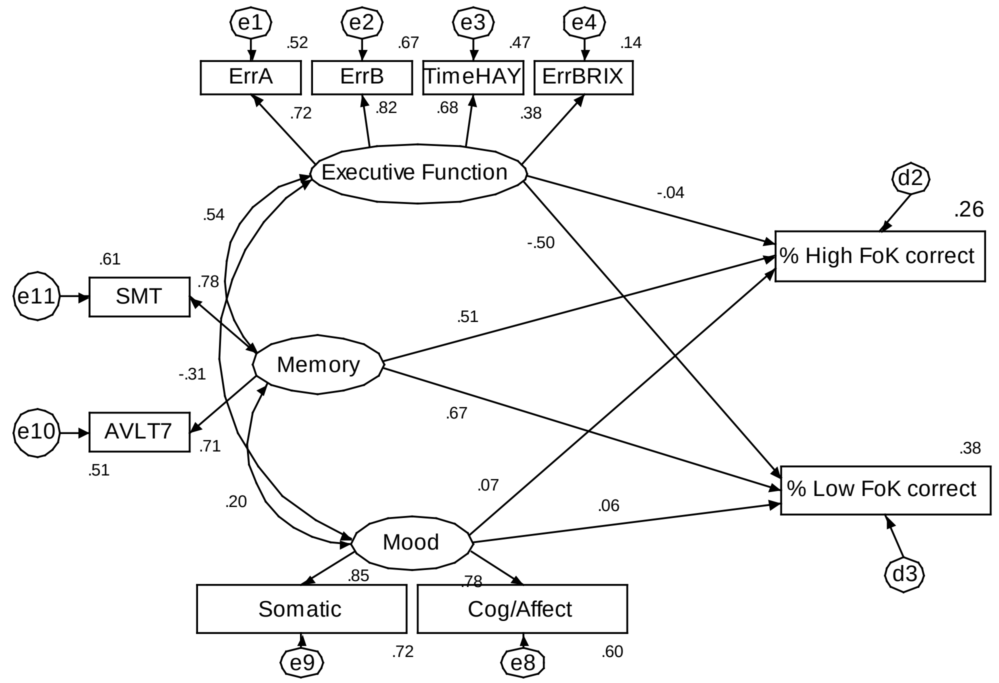
	RCJgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJgamma	1.000								
SMTrecall	.341	1.000							
AVLT7	.138	.569	1.000						
Somatic	.030	.129	.071	1.000					
Cog/Affect	.123	.109	.098	.625	1.000				
ErrBRIX	.047	.255	.380	-.142	-.079	1.000			
TimeHAY	-.014	.186	.267	-.141	-.127	.272	1.000		
ErrB	.016	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	.035	.337	.383	-.226	-.116	.244	.445	.518	1.000

Residual Covariances

	RCJgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJgamma	.005								
SMTrecall	.093	2.864							
AVLT7	-.056	1.434	2.252						
Somatic	.000	.079	-.102	.021					
Cog/Affect	.019	.135	.097	.003	.023				
ErrBRIX	.016	2.503	5.288	-.189	-.029	5.823			
TimeHAY	-.787	-20.230	-.270	2.378	.228	7.077	174.601		
ErrB	-.024	-2.075	-.541	-.088	-.046	-.707	11.919	.702	
ErrA	-.009	.196	.471	-.022	.035	-.234	-3.011	-.009	.083

Standardized Residual Covariances

	RCJgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
RCJgamma	1.066								
SMTrecall	1.329	1.482							
AVLT7	-.877	.998	1.395						
Somatic	-.025	.244	-.344	.200					
Cog/Affect	1.091	.392	.310	.026	.187				
ErrBRIX	.152	1.157	2.680	-.376	-.055	1.245			
TimeHAY	-.827	-1.008	-.015	.518	.047	.229	.461		
ErrB	-.563	-2.284	-.655	-.427	-.212	-.506	.878	.922	
ErrA	-.343	.370	.980	-.186	.275	-.288	-.383	-.024	.319



The model is recursive.
Sample size = 100

Variable Summary

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FoK accuracy	% High FoK correct
FoK inaccuracy	% Low FoK correct

Unobserved, endogenous variables

Info Process

Unobserved, exogenous variables

Executive Function
Mood
Memory

Variable counts

Number of variables in your model:	23
Number of observed variables:	10
Number of unobserved variables:	13
Number of exogenous variables:	13
Number of endogenous variables:	10

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	13	0	0	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	11	3	13	0	0	27
Total	24	3	13	0	0	40

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	27
Degrees of freedom (55 - 27):	28

Result

Minimum was achieved
Chi-square = 35.482
Degrees of freedom = 28
Probability level = .156

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	27	35.482	28	.156	1.267
Saturated model	55	.000	0		
Independence model	10	109.856	45	.000	2.441
Zero model	0	495.000	55	.000	9.000

RMR, GFI

Model	GFI
Default model	.928
Saturated model	1.000
Independence model	.778
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.885
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.052	.000	.099	.442
Independence model	.121	.092	.149	.000

Regression Weights

			Estimate	S.E.	C.R.	P
ErrA	<---	Executive Function	.027	.006	4.583	***
ErrB	<---	Executive Function	.054	.010	5.215	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.060	.026	2.301	.021
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.019	.304	3.354	***
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	1.213	.297	4.077	***
FoK accuracy	<---	Executive Function	-.029	.137	-.209	.834
FoK accuracy	<---	Memory	5.311	2.484	2.138	.032
FoK accuracy	<---	Mood	2.492	5.063	.492	.623
FoK inaccuracy	<---	Executive Function	-.355	.169	-2.100	.036
FoK inaccuracy	<---	Memory	7.228	2.382	3.034	.002
FoK inaccuracy	<---	Mood	2.163	5.972	.362	.717

Standardized Regression Weights

			Estimate
ErrA	<---	Executive Function	.719
ErrB	<---	Executive Function	.821
TimeHAY	<---	Executive Function	.682
ErrBRIX	<---	Executive Function	.376
Cog/Affect	<---	Mood	.776
Somatic	<---	Mood	.847
AVLT7	<---	Memory	.712
SMTrecall	<---	Memory	.782
FoK accuracy	<---	Executive Function	-.041
FoK accuracy	<---	Memory	.505
FoK accuracy	<---	Mood	.073
FoK inaccuracy	<---	Executive Function	-.496
FoK inaccuracy	<---	Memory	.670
FoK inaccuracy	<---	Mood	.062

Covariances:

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	44.640	24.525	1.820	.069
Mood	<-->	Memory	.342	.257	1.327	.184
Executive Function	<-->	Mood	-7.952	3.463	-2.297	.022

Correlations:

			Estimate
Executive Function	<-->	Memory	.540
Mood	<-->	Memory	.203
Executive Function	<-->	Mood	-.313

Variances:

	Estimate	S.E.	C.R.	P
Executive Function	1244.376	395.357	3.147	.002
Mood	.518	.189	2.734	.006
Memory	5.488	2.336	2.349	.019
e1	.835	.173	4.823	***
e2	1.740	.574	3.030	.002
e3	1429.547	271.369	5.268	***
e4	27.044	4.437	6.096	***
e8	.341	.139	2.462	.014
e9	.212	.137	1.543	.123
e1o	5.349	1.200	4.457	***
e11	5.113	1.446	3.535	***
d2	450.889	78.169	5.768	***
d3	399.143	83.545	4.778	***

Squared Multiple Correlations:

	Estimate
FoK inaccuracy	.375
FoK accuracy	.257
SMTrecall	.612
AVLT7	.506
Somatic	.717
Cog/Affect	.603
ErrBRIX	.142
TimeHAY	.465
ErrB	.674
ErrA	.517

Sample Moments

Sample Covariances

	FoK inaccuracy	FoK accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	767.137									
FoK accuracy	173.177	674.967								
SMTrecall	46.325	46.132	16.462							
AVLT7	35.002	22.714	8.511	13.608						
Somatic	5.769	3.737	.459	.230	.774					
Cog/Affect	8.427	3.087	.414	.341	.516	.881				
ErrBRIX	12.087	24.190	6.441	8.723	-.775	-.459	38.735			
TimeHAY	-44.481	139.842	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	-15.690	5.191	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	-.109	2.190	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	FoK inaccuracy	FoK accuracy	SMTrecall	AVLT7	Somatic	Cog/Affectχ	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	1.000									
FoK accuracy	.241	1.000								
SMTrecall	.412	.438	1.000							
AVLT7	.343	.237	.569	1.000						
Somatic	.237	.164	.129	.071	1.000					
Cog/Affect	.324	.127	.109	.098	.625	1.000				
ErrBRIX	.070	.150	.255	.380	-.142	-.079	1.000			
TimeHAY	-.030	.101	.186	.267	-.141	-.127	.272	1.000		
ErrB	-.230	.081	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	-.003	.061	.337	.383	-.226	-.116	.244	.445	.518	1.000

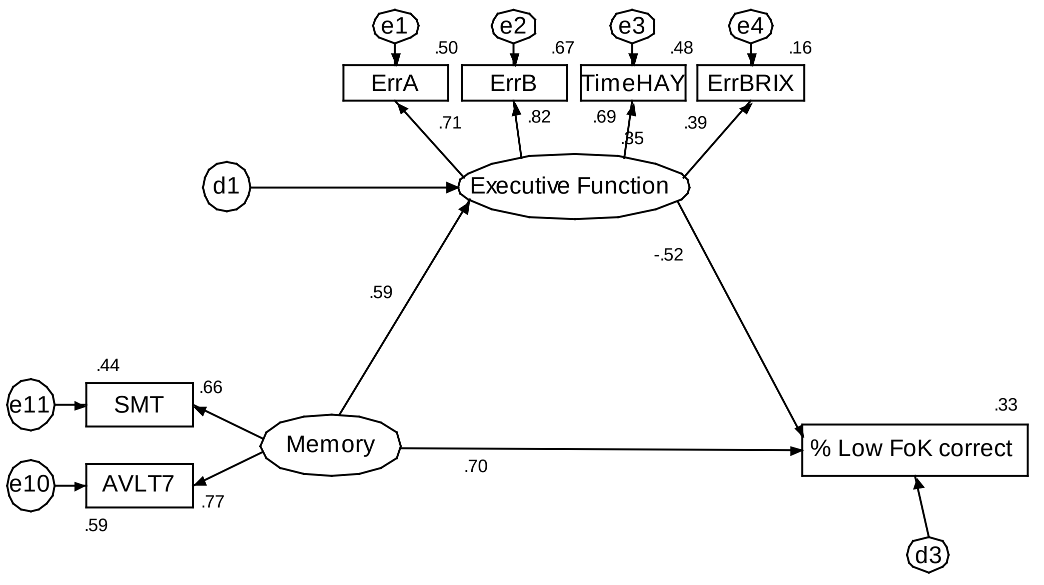
Residual Covariances

	FoK inaccarucy	FoK accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccarucy	128.161									
FoK accuracy	22.873	68.436								
SMTrecall	16.550	11.302	3.278							
AVLT7	10.448	-6.007	1.855	2.770						
Somatic	-.764	.343	.037	-.118	.025					
sqCAbdi1	2.013	-.245	.000	-.001	-.011	.023				
ErrBRIX	20.264	13.312	3.199	6.049	-.290	.018	7.226			
TimeHAY	92.027	-41.765	-13.937	7.787	1.494	1.589	15.538	164.505		
ErrB	-8.356	-4.567	-1.992	-.331	-.098	.055	-.603	8.107	.722	
ErrA	3.549	-2.676	.439	.752	-.057	.063	.095	-.652	-.033	.176

Standardized Residual Covariances

	FoK inaccarucy	FoK accuracy	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccarucy	1.411									
FoK accuracy	.355	.794								
SMTrecall	1.707	1.172	1.749							
AVLT7	1.198	-.695	1.349	1.799						
Somatic	-.333	.158	.117	-.408	.236					
Cog/Affect	.825	-.106	.000	-.003	-.112	.186				
ErrBRIX	1.419	.955	1.542	3.223	-.590	.033	1.613			
TimeHAY	.697	-.323	-.710	.440	.327	.326	.516	.433		
ErrB	-1.413	-.788	-2.233	-.413	-.476	.250	-.442	.589	.952	
ErrA	1.056	-.813	.876	1.667	-.491	.509	.124	-.086	-.092	.718

APPENDIX S - FEELING OF KNOWING INACCURATE CALIBRATION



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FoK inaccuracy	% Low FoK correct

Unobserved, endogenous variables

Executive Function

Unobserved, exogenous variables

Memory

Variable counts (including residuals)

Number of variables in your model:	17
Number of observed variables:	7
Number of unobserved variables:	10
Number of exogenous variables:	9
Number of endogenous variables:	8

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	10	0	0	0	0	10
Labeled	0	0	0	0	0	0
Unlabeled	7	0	9	0	0	16
Total	17	0	9	0	0	26

Computation of degrees of freedom

Number of distinct sample moments:	28
Number of distinct parameters to be estimated:	16
Degrees of freedom (28 - 16):	12

Result

Minimum was achieved
Chi-square = 19.439
Degrees of freedom = 12
Probability level = .078

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	16	19.439	12	.078	1.620
Saturated model	28	.000	0		
Independence model	7	65.668	21	.000	3.127
Zero model	0	346.500	28	.000	12.375

RMR, GFI

Model	GFI
Default model	.944
Saturated model	1.000
Independence model	.810
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.833
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.079	.000	.141	.209
Independence model	.147	.107	.187	.000

Regression Weights

			Estimate	S.E.	C.R.	P
Executive Function	<---	Memory	8.313	2.517	3.302	***
ErrA	<---	Executive Function	.027	.005	4.985	***
ErrB	<---	Executive Function	.053	.010	5.548	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.063	.023	2.711	.007
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.938	.211	4.437	***
FoK inaccuracy	<---	Executive Function	-.367	.146	-2.512	.012
FoK inaccuracy	<---	Memory	6.923	2.340	2.958	.003

Standardized Regression Weights

			Estimate
Executive Function	<---	Memory	.591
ErrA	<---	Executive Function	.710
ErrB	<---	Executive Function	.820
TimeHAY	<---	Executive Function	.692
ErrBRIX	<---	Executive Function	.394
AVLT7	<---	Memory	.765
SMTrecall	<---	Memory	.664
FoK inaccuracy	<---	Executive Function	-.524
FoK inaccuracy	<---	Memory	.704

Variances

	Estimate	S.E.	C.R.	P
Memory	6.680	2.292	2.915	.004
d1	859.197	350.107	2.454	.014
e1	.914	.183	5.008	***
e2	1.828	.589	3.101	.002
e3	1441.012	276.062	5.220	***
e4	28.479	4.564	6.240	***
e1o	4.721	1.418	3.329	***
e11	7.457	1.545	4.827	***
d3	429.756	94.264	4.559	***

Squared Multiple Correlations

	Estimate
Executive Function	.350
FoK inaccuracy	.334
SMTrecall	.441
AVLT7	.586
ErrBRIX	.155
TimeHAY	.478
ErrB	.672
ErrA	.505

Sample Moments

Sample Covariances

	FoK inaccuracy	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	767.137						
SMTrecall	46.325	16.462					
AVLT7	35.002	8.511	13.608				
ErrBRIX	12.087	6.441	8.723	38.735			
TimeHAY	-44.481	40.196	52.428	90.076	2838.428		
ErrB	-15.690	.917	2.068	3.402	74.968	6.054	
ErrA	-.109	1.889	1.949	2.092	32.689	1.759	1.905

Sample Correlations

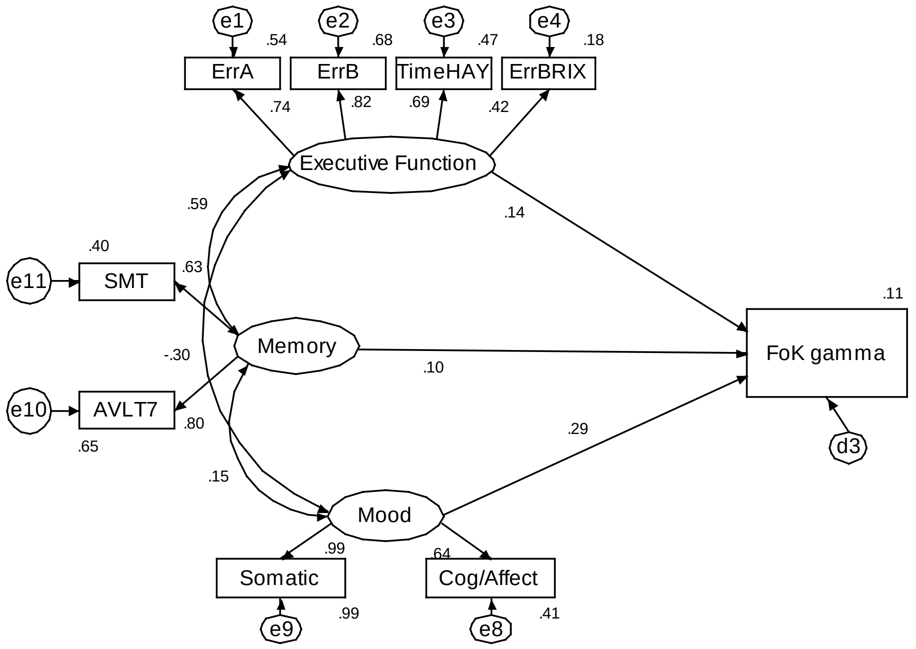
	FoK inaccuracy	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	1.000						
SMTrecall	.412	1.000					
AVLT7	.343	.569	1.000				
ErrBRIX	.070	.255	.380	1.000			
TimeHAY	-.030	.186	.267	.272	1.000		
ErrB	-.230	.092	.228	.222	.572	1.000	
ErrA	-.003	.337	.383	.244	.445	.518	1.000

Residual Covariances

	FoK inaccuracy	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	121.555						
SMTrecall	22.048	3.130					
AVLT7	9.114	2.246	2.207				
ErrBRIX	18.375	3.160	5.224	5.014			
TimeHAY	55.331	-11.883	-3.107	6.862	76.559		
ErrB	-10.379	-1.854	-.887	-1.025	4.687	.487	
ErrA	2.542	.506	.474	-.117	-2.387	-.108	.059

Standardized Residual Covariances

	FoK inaccuracy	SMTrecall	AVLT7	ErrBRIX	TimeHAY	ErrB	ErrA
FoK inaccuracy	1.325						
SMTrecall	2.288	1.652					
AVLT7	1.012	1.616	1.362				
ErrBRIX	1.238	1.466	2.610	1.046			
TimeHAY	.411	-.595	-.166	.216	.195		
ErrB	-1.716	-2.039	-1.039	-.709	.327	.616	
ErrA	.730	.979	.978	-.142	-.299	-.288	.226



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FoK	FoK Gamma

Unobserved, exogenous variables
Executive Function
Mood
Memory

Variable counts

Number of variables in your model:	21
Number of observed variables:	9
Number of unobserved variables:	12
Number of exogenous variables:	12
Number of endogenous variables:	9

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	12	0	1	0	0	13
Labeled	0	0	0	0	0	0
Unlabeled	8	3	11	0	0	22
Total	20	3	12	0	0	35

Computation of degrees of freedom

Number of distinct sample moments:	45
Number of distinct parameters to be estimated:	22
Degrees of freedom (45 - 22):	23

Result (Default model)

Minimum was achieved
Chi-square = 25.483
Degrees of freedom = 23
Probability level = .326

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	22	25.483	23	.326	1.108
Saturated model	45	.000	0		
Independence model	9	99.212	36	.000	2.756
Zero model	0	445.500	45	.000	9.900

RMR, GFI

Model	GFI
Default model	.943
Saturated model	1.000
Independence model	.777
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.961
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.033	.000	.091	.620
Independence model	.133	.102	.165	.000

Regression Weights

			Estimate	S.E.	C.R.	P
ErrA	<---	Executive Function	.027	.006	4.909	***
ErrB	<---	Executive Function	.054	.010	5.642	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.067	.024	2.804	.005
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.446	.186	7.765	***
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.861	.218	3.943	***
FoKgamma	<---	Executive Function	.001	.002	.641	.522
FoKgamma	<---	Memory	.013	.027	.499	.618
FoKgamma	<---	Mood	.168	.087	1.947	.052

Standardized Regression Weights

			Estimate
ErrA	<---	Executive Function	.736
ErrB	<---	Executive Function	.824
TimeHAY	<---	Executive Function	.688
ErrBRIX	<---	Executive Function	.418
Cog/Affect	<---	Mood	.641
Somatic	<---	Mood	.993
AVLT7	<---	Memory	.804
SMTrecall	<---	Memory	.634
FoKgamma	<---	Executive Function	.139
FoKgamma	<---	Memory	.103
FoKgamma	<---	Mood	.286

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Memory	57.837	20.515	2.819	.005
Mood	<-->	Memory	.239	.226	1.058	.290
Executive Function	<-->	Mood	-6.381	2.674	-2.387	.017

Correlations

			Estimate
Executive Function	<-->	Memory	.588
Mood	<-->	Memory	.147
Executive Function	<-->	Mood	-.297

Variances

	Estimate	S.E.	C.R.	P
Executive Function	1302.667	387.502	3.362	***
Mood	.354	.103	3.421	***
Memory	7.427	2.504	2.966	.003
e9	.010			
e1	.825	.176	4.697	***
e2	1.768	.576	3.070	.002
e3	1448.941	270.848	5.350	***
e4	27.421	4.469	6.136	***
e8	.507	.074	6.811	***
e1o	4.058	1.766	2.299	.022
e11	8.187	1.752	4.674	***
d3	.109	.017	6.423	***

Squared Multiple Correlations

	Estimate
FoKgamma	.114
SMTrecall	.402
AVLT7	.647
Somatic	.987
Cog/Affect	.411
ErrBRIX	.175
TimeHAY	.473
ErrB	.680
ErrA	.541

Sample Moments

Sample Covariances

	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoKgamma	.139								
SMTrecall	.068	16.462							
AVLT7	.099	8.511	13.608						
Somatic	.078	.459	.230	.774					
Cog/Affect	.021	.414	.341	.516	.881				
ErrBRIX	.112	6.441	8.723	-.775	-.459	38.735			
TimeHAY	1.372	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	.169	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	-.020	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoKgamma	1.000								
SMTrecall	.045	1.000							
AVLT7	.072	.569	1.000						
Somatic	.239	.129	.071	1.000					
Cog/Affect	.061	.109	.098	.625	1.000				
ErrBRIX	.048	.255	.380	-.142	-.079	1.000			
TimeHAY	.069	.186	.267	-.141	-.127	.272	1.000		
ErrB	.184	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	-.040	.337	.383	-.226	-.116	.244	.445	.518	1.000

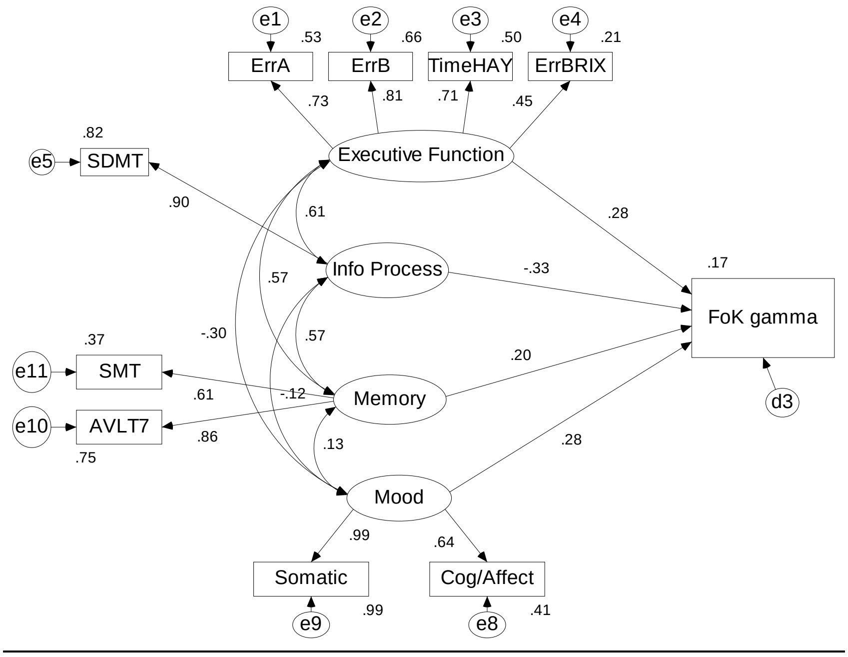
Residual Covariances

	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoKgamma	.016								
SMTrecall	-.119	2.763							
AVLT7	-.118	2.113	2.122						
Somatic	.000	.162	-.115	.024					
Cog/Affect	-.033	.208	.102	.005	.021				
ErrBRIX	.015	3.111	4.857	-.158	-.032	5.495			
TimeHAY	-.080	-9.626	-5.409	2.622	.019	3.010	86.819		
ErrB	.091	-1.756	-1.035	-.038	-.030	-1.269	5.080	.537	
ErrA	-.060	.528	.368	-.022	.025	-.287	-2.911	-.151	.107

Standardized Residual Covariances

	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
FoKgamma	.917								
SMTrecall	-.905	1.419							
AVLT7	-.971	1.493	1.300						
Somatic	-.001	.500	-.388	.225					
Cog/Affect	-.987	.603	.322	.051	.170				
ErrBRIX	.074	1.433	2.426	-.313	-.060	1.163			
TimeHAY	-.043	-.478	-.288	.563	.004	.095	.222		
ErrB	1.095	-1.921	-1.205	-.180	-.136	-.881	.357	.685	
ErrA	-1.269	1.021	.761	-.185	.195	-.353	-.367	-.408	.419

FEELING OF KNOWING GAMMA, DIRECT EFFECTS, INCLUDING INFORMATION PROCESSING



The model is recursive.
Sample size = 100

Variable Summary

Your model contains the following variables (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FoK	FoK Gamma
SDMT	Symbol Digit Modalities test

Unobserved, exogenous variables

Executive Function
Mood
Memory
Info Process

Variable counts (including residuals)

Number of variables in your model:	24
Number of observed variables:	10
Number of unobserved variables:	14
Number of exogenous variables:	14
Number of endogenous variables:	10

Parameter summary (Group number 1)

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	14	0	2	0	0	16
Labeled	0	0	0	0	0	0
Unlabeled	9	6	12	0	0	27
Total	23	6	14	0	0	43

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	27
Degrees of freedom (55 - 27):	28

Result (Default model)

Minimum was achieved
Chi-square = 29.286
Degrees of freedom = 28
Probability level = .398

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	27	29.286	28	.398	1.046
Saturated model	55	.000	0		
Independence model	10	109.769	45	.000	2.439
Zero model	0	495.000	55	.000	9.000

RMR, GFI

Model	GFI
Default model	.941
Saturated model	1.000
Independence model	.778
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.980
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.022	.000	.082	.714
Independence model	.121	.092	.149	.000

Regression Weights:

			Estimate	S.E.	C.R.	P
ErrA	<---	Executive Function	.026	.005	4.906	***
ErrB	<---	Executive Function	.051	.009	5.772	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.071	.024	2.941	.003
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.445	.187	7.732	***
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.760	.189	4.024	***
SDMT	<---	Info Process	1.000			
FoKgamma	<---	Executive Function	.003	.002	1.325	.185
FoKgamma	<---	Info Process	-.012	.006	-1.970	.049
FoKgamma	<---	Memory	.025	.024	1.029	.303
FoKgamma	<---	Mood	.163	.083	1.971	.049

Standardized Regression Weights:

			Estimate
ErrA	<---	Executive Function	.731
ErrB	<---	Executive Function	.811
TimeHAY	<---	Executive Function	.709
ErrBRIX	<---	Executive Function	.455
Cog/Affect	<---	Mood	.640
Somatic	<---	Mood	.993
AVLT7	<---	Memory	.864
SMTrecall	<---	Memory	.605
SDMT	<---	Info Process	.893
FoKgamma	<---	Executive Function	.288
FoKgamma	<---	Info Process	-.344
FoKgamma	<---	Memory	.209
FoKgamma	<---	Mood	.277

Covariances:

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Info Process	237.591	71.332	3.331	***
Memory	<-->	Info Process	17.605	5.727	3.074	.002
Mood	<-->	Memory	.221	.230	.959	.338
Mood	<-->	Info Process	-.766	.752	-1.018	.309
Executive Function	<-->	Mood	-6.645	2.747	-2.419	.016
Executive Function	<-->	Memory	62.654	19.978	3.136	.002

Correlations:

			Estimate
Executive Function	<-->	Info Process	.616
Memory	<-->	Info Process	.573
Mood	<-->	Memory	.126
Mood	<-->	Info Process	-.124
Executive Function	<-->	Mood	-.303
Executive Function	<-->	Memory	.574

Variances:

	Estimate	S.E.	C.R.	P
Executive Function	1369.756	399.325	3.430	***
Mood	.352	.103	3.409	***
Memory	8.695	2.565	3.389	***
Info Process	108.486	23.034	4.710	***
e9	.010			
e5	27.703			
e1	.824	.172	4.799	***
e2	1.890	.540	3.502	***
e3	1355.570	258.416	5.246	***
e4	26.161	4.345	6.021	***
e8	.507	.074	6.814	***
e1o	2.946	1.753	1.680	.093
e11	8.692	1.651	5.264	***
d3	.101	.017	6.076	***

Squared Multiple Correlations:

	Estimate
SDMT	.797
FoKgamma	.176
SMTrecall	.366
AVLT7	.747
Somatic	.987
Cog/Affect	.410
ErrBRIX	.207
TimeHAY	.503
ErrB	.657
ErrA	.534

Sample Moments

Sample Covariances

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	152.365									
FoKgamma	-.548	.139								
SMTrecall	15.103	.068	16.462							
AVLT7	22.784	.099	8.511	13.608						
Somatic	-1.132	.078	.459	.230	.774					
Cog/Affect	-.493	.021	.414	.341	.516	.881				
ErrBRIX	29.043	.112	6.441	8.723	-.775	-.459	38.735			
TimeHAY	257.314	1.372	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	8.959	.169	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	6.160	-.020	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	1.000									
FoKgamma	-.119	1.000								
SMTrecall	.302	.045	1.000							
AVLT7	.500	.072	.569	1.000						
Somatic	-.104	.239	.129	.071	1.000					
Cog/Affect	-.043	.061	.109	.098	.625	1.000				
ErrBRIX	.378	.048	.255	.380	-.142	-.079	1.000			
TimeHAY	.391	.069	.186	.267	-.141	-.127	.272	1.000		
ErrB	.295	.184	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	.362	-.040	.337	.383	-.226	-.116	.244	.445	.518	1.000

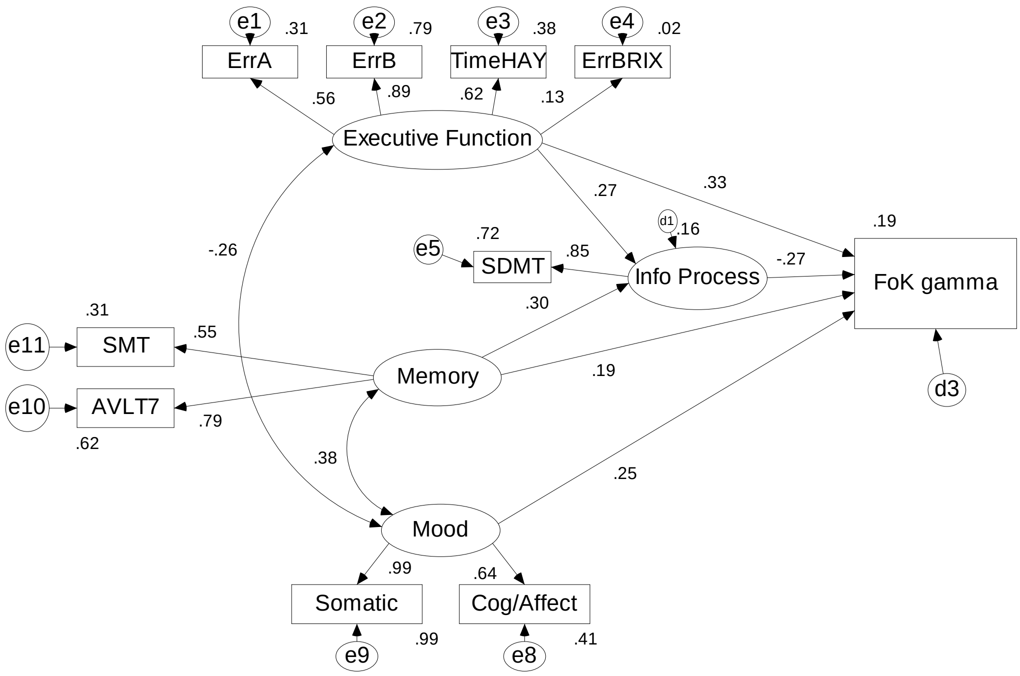
Residual Covariances (Group number 1 - Default model)

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	16.176									
FoKgamma	-.254	.017								
SMTrecall	1.725	-.098	2.748							
AVLT7	5.178	-.119	1.903	1.966						
Somatic	-.024	.001	.217	-.089	.028					
Cog/Affect	.273	-.032	.246	.120	.008	.022				
ErrBRIX	12.269	.009	3.080	4.299	-.097	.010	5.747			
TimeHAY	19.723	-.082	-7.415	-10.227	2.999	.282	-6.628	113.102		
ErrB	-3.259	.094	-1.532	-1.154	-.039	-.031	-1.571	4.529	.542	
ErrA	-.079	-.059	.639	.303	-.022	.025	-.447	-3.283	-.091	.136

Standardized Residual Covariances

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	.836									
FoKgamma	-.617	.952								
SMTrecall	.379	-.750	1.410							
AVLT7	1.183	-.979	1.328	1.188						
Somatic	-.024	.024	.672	-.299	.267					
Cog/Affect	.250	-.974	.713	.376	.080	.184				
ErrBRIX	1.767	.047	1.423	2.129	-.192	.019	1.226			
TimeHAY	.300	-.044	-.371	-.539	.648	.058	-.209	.292		
ErrB	-1.081	1.137	-1.687	-1.330	-.187	-.139	-1.087	.319	.692	
ErrA	-.047	-1.250	1.251	.625	-.186	.197	-.553	-.418	-.249	.542

APPENDIX V - FoK gamma Mediation Model



The model is recursive.
Sample size = 100

Variable Summary (residuals not shown)

Observed, endogenous variables

ErrA	A type Errors, Hayling Test
ErrB	B type errors, Hayling Test
TimeHAY	Hayling Time (complex minus simple)
ErrBRIX	Errors Brixton Test
Cog/Affect	BDI-II Cognitive Affective Factor
Somatic	BDI-II Somatic Factor
AVLT7	Auditory Verbal Learning Test, delayed recall.
SMTrecall	Sentence Memory Test recall
FoK	FoK Gamma
SDMT	Symbol Digit Modalities Test

Unobserved, endogenous variables
Info Process

Unobserved, exogenous variables
Executive Function
Mood
Memory

Variable counts (including residuals)

Number of variables in your model:	25
Number of observed variables:	10
Number of unobserved variables:	15
Number of exogenous variables:	14
Number of endogenous variables:	11

Parameter summary

	Weights	Covariances	Variances	Means	Intercepts	Total
Fixed	15	0	2	0	0	17
Labeled	0	0	0	0	0	0
Unlabeled	11	2	12	0	0	25
Total	26	2	14	0	0	42

Computation of degrees of freedom

Number of distinct sample moments:	55
Number of distinct parameters to be estimated:	25
Degrees of freedom (55 - 25):	30

Result

Minimum was achieved
Chi-square = 37.786
Degrees of freedom = 30
Probability level = 0.155

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	25	37.786	30	.155	1.260
Saturated model	55	.000	0		
Independence model	10	109.769	45	.000	2.439
Zero model	0	495.000	55	.000	9.000

RMR, GFI

Model	GFI
Default model	.924
Saturated model	1.000
Independence model	.778
Zero model	.000

Baseline Comparisons

Model	CFI
Default model	.880
Saturated model	1.000
Independence model	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.051	.000	.097	.452
Independence model	.121	.092	.149	.000

Regression Weights

			Estimate	S.E.	C.R.	P
Info Process	<---	Executive Function	.076	.044	1.727	.084
Info Process	<---	Memory	1.106	.625	1.770	.077
ErrA	<---	Executive Function	.021	.005	3.961	***
ErrB	<---	Executive Function	.070	.019	3.607	***
TimeHAY	<---	Executive Function	1.000			
ErrBRIX	<---	Executive Function	.022	.023	.950	.342
Cog/Affect	<---	Mood	1.000			
Somatic	<---	Mood	1.419	.181	7.840	***
AVLT7	<---	Memory	1.000			
SMTrecall	<---	Memory	.854	.370	2.306	.021
SDMT	<---	Info Process	1.000			
FoKgamma	<---	Mood	.148	.081	1.820	.069
FoKgamma	<---	Executive Function	.004	.002	2.323	.020
FoKgamma	<---	Memory	.030	.027	1.105	.269
FoKgamma	<---	Info Process	-.011	.006	-1.910	.056

Standardized Regression Weights

			Estimate
Info Process	<---	Executive Function	.270
Info Process	<---	Memory	.298
ErrA	<---	Executive Function	.561
ErrB	<---	Executive Function	.891
TimeHAY	<---	Executive Function	.619
ErrBRIX	<---	Executive Function	.128
Cog/Affect	<---	Mood	.644
Somatic	<---	Mood	.993
AVLT7	<---	Memory	.786
SMTrecall	<---	Memory	.553
SDMT	<---	Info Process	.847
FoKgamma	<---	Mood	.249
FoKgamma	<---	Executive Function	.331
FoKgamma	<---	Memory	.192
FoKgamma	<---	Info Process	-.265

Covariances

			Estimate	S.E.	C.R.	P
Executive Function	<-->	Mood	-4.700	2.815	-1.670	.095
Mood	<-->	Memory	.519	.214	2.426	.015

Correlations

			Estimate
Executive Function	<-->	Mood	-.262
Mood	<-->	Memory	.383

Variances

	Estimate	S.E.	C.R.	P
Executive Function	900.605	387.531	2.324	.020
Mood	.358	.103	3.465	***
Memory	5.126	2.560	2.002	.045
d1	59.109	14.279	4.139	***
e9	.010			
e5	27.703			
e1	.898	.168	5.352	***
e2	1.137	.905	1.257	.209
e3	1451.299	273.806	5.300	***
e4	26.488	4.312	6.142	***
e8	.506	.074	6.806	***
e10	3.170	1.989	1.594	.111
e11	8.492	1.896	4.478	***
d3	.102	.017	5.992	***

Squared Multiple Correlations

	Estimate
Info Process	.162
SDMT	.718
FoKgamma	.189
SMTrecall	.306
AVLT7	.618
Somatic	.986
Cog/Affect	.414
ErrBRIX	.017
TimeHAY	.383
ErrB	.794
ErrA	.314

Sample Moments

Sample Covariances

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	152.365									
FoKgamma	-.548	.139								
SMTrecall	15.103	.068	16.462							
AVLT7	22.784	.099	8.511	13.608						
Somatic	-1.132	.078	.459	.230	.774					
Cog/Affect	-.493	.021	.414	.341	.516	.881				
ErrBRIX	29.043	.112	6.441	8.723	-.775	-.459	38.735			
TimeHAY	257.314	1.372	40.196	52.428	-6.606	-6.363	90.076	2838.428		
ErrB	8.959	.169	.917	2.068	-.533	-.372	3.402	74.968	6.054	
ErrA	6.160	-.020	1.889	1.949	-.274	-.150	2.092	32.689	1.759	1.905

Sample Correlations

	SDMT	SPSSgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	1.000									
FoKgamma	-.119	1.000								
SMTrecall	.302	.045	1.000							
AVLT7	.500	.072	.569	1.000						
Somatic	-.104	.239	.129	.071	1.000					
Cog/Affect	-.043	.061	.109	.098	.625	1.000				
ErrBRIX	.378	.048	.255	.380	-.142	-.079	1.000			
TimeHAY	.391	.069	.186	.267	-.141	-.127	.272	1.000		
ErrB	.295	.184	.092	.228	-.246	-.161	.222	.572	1.000	
ErrA	.362	-.040	.337	.383	-.226	-.116	.244	.445	.518	1.000

Residual Covariances

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	54.126									
FoKgamma	-.227	.013								
SMTrecall	10.263	-.075	4.232							
AVLT7	17.115	-.068	4.134	5.312						
Somatic	-1.441	.011	-.169	-.506	.043					
Cog/Affect	-.711	-.026	-.029	-.178	.008	.017				
ErrBRIX	27.529	.066	6.441	8.723	-.627	-.354	11.802			
TimeHAY	189.160	-.696	40.196	52.428	.062	-1.663	70.063	486.524		
ErrB	4.206	.025	.917	2.068	-.068	-.045	2.007	12.158	.537	
ErrA	4.703	-.065	1.889	1.949	-.132	-.049	1.665	13.432	.416	.595

Standardized Residual Covariances

	SDMT	FoKgamma	SMTrecall	AVLT7	Somatic	Cog/Affect	ErrBRIX	TimeHAY	ErrB	ErrA
SDMT	3.876									
FoKgamma	-.638	.717								
SMTrecall	2.918	-.597	2.435							
AVLT7	5.851	-.652	3.746	4.505						
Somatic	-1.691	.339	-.551	-1.958	.411					
Cog/Affect	-.767	-.783	-.087	-.649	.087	.136				
ErrBRIX	5.323	.357	3.531	5.806	-1.404	-.730	3.083			
TimeHAY	3.877	-.399	2.358	3.735	.015	-.365	2.761	1.455		
ErrB	1.761	.291	1.110	3.041	-.328	-.201	1.627	.930	.685	
ErrA	4.091	-1.575	4.697	5.881	-1.325	-.459	2.781	2.275	1.377	3.198