UNIVERSITY OF SOUTHERN QUEENSLAND

COMPETING TASKS AS MEASURES OF INTELLIGENCE AND PREDICTORS OF JOB PERFORMANCE

A Dissertation submitted by

Bradley Dolph

For the award of

Doctor of Philosophy

2012

Abstract

This series of studies investigated a new measure of cognitive ability, the Multi-Tasks test, its place within the structure of intelligence and its usefulness in predicting job performance. The Multi-Tasks test employed a competing task methodology, being the simultaneous performance of two cognitive tasks, which has been shown to have a significant relationship with intelligence and job performance, particularly for complex jobs. The competing tasks methodology has a long history in psychology research and has recently experienced a resurgence of interest as technological advances (e.g., the Internet) have made it easier to administer these measures within the workplace. In the pilot study (Part A of Study 1) the means, reliability and demographic group differences of the measure were investigated. In Part B of Study 1 and Studies 2 and 3, the reliability and predictive validity of the test was compared to measures of general mental ability (crystallized and fluid intelligence) which have been widely used in personnel selection. Crystallized intelligence measures are language based and influenced by culture and education, whereas fluid intelligence tasks typically draw on non-verbal reasoning and are unaffected by education. These measures feature prominently in the Cattell-Horn-Carroll Theory of Cognitive Abilities, which forms the theoretical basis for these studies. In Study 2 and Study 3, additional cognitive measures were added to further elucidate the place of Multi-Tasks within the intelligence model, including a measure of short-term memory (Gsm in the CHC Theory). Previous research shows short-term memory and a related concept working memory, to be important in performance on the Multi-Tasks test. Further, the reliability and predictive validity of Multi-Tasks was compared to a personality measure (the Big Five model of personality) in Study 2, which is also widely used in job selection.

In all studies the Multi-Tasks test had high reliability, and it was found to be a more reliable measure than the general mental ability measures in Study 1 (Part B), Study 2 and

Study 3. In Study 1 (Part B) it was more highly correlated with the fluid than the crystallized intelligence measure. The addition of the short-term memory task in Study 2 revealed that the highest correlation was between Multi-Tasks and Gsm, however this factor did not appear in Study 3 and Multi-Tasks was, as per Study 1, a Gf measure. These findings support previous research demonstrating that the measure is likely to be relatively independent of the influence of culture and language and that it draws on working memory ability. All studies showed Multi-Tasks to be a good predictor of job performance. It strongly predicted two of three measures of job performance in Study 1 (Part B), three of four measures in Study 2 and it was positively associated with 1 out of 3 job performance indicators in Study 3. The other cognitive measures also predicted some measures of job performance in all studies, but not as strongly or consistently as Multi-Tasks. Study 3 demonstrated that the factor structure and reliability of the measure in a sample of Chinese workers was comparable to previous studies, which indicates that the measure is not affected by culture and can be employed cross-culturally. Other group differences in performance on the Multi-Tasks test were not consistent between studies, however where they did exist they showed older and more highly educated workers to perform better. This supports research showing that the Multi-Tasks test shows promise as a predictor of performance in complex jobs and managerial potential.

There were a number of limitations discussed and many opportunities for further research. Overall the results of these studies indicate that the Multi-Tasks test shows promise as a valid, reliable, culturally unbiased measure of job performance that is suitable for a variety of job roles, both in Australia and cross-culturally, and may be particularly useful as an indicator of management potential. As a new test, further research to replicate these findings is encouraged.

Keywords: competing tasks, multi-tasks, CHC Theory, attention, working memory, job performance, culture, culturally unbiased.

CERTIFICATION OF DISSERTATION

I certify that the ideas, experimental work, results analyses, software and conclusions reported in this dissertation are entirely my own effort, expect where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, expect where otherwise acknowledged.

Signature of Candidate Date

ENDORSEMENT

Signature of Supervisor/s Date

ACKNOWLEDGEMENTS

First and foremost I want to thank my principal supervisor, Professor Gerard Fogarty for your unwavering support and guidance throughout my PhD journey. Finding you was a Godsend.

To my associate supervisor Professor Lazar Stankov whose career inspired me to begin my research. Thank you for all your encouragement over the last decade.

In the early days of my dissertation and planning, I must give a special thanks to my friends Dr Richard Roberts and Dr John Crawford for their enthusiasm and guidance.

A special thanks to my programmer and Web wizard Giles Wycherley. There is nothing you cannot do when it comes to software.

Many thanks to those organisations who helped provide the data for this research program. Without you, this PhD would not be possible. Specifically, I would like to thank HAYS, Bankwest and Goodyear for their support.

I would like to dedicate this PhD to my son Toby and wife Melinda, the most important people in my life. To my Mum and Dad who taught me to never give up and encouraged me to follow my dreams. I have been truly blessed.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER 1

INTRODUCTION

COMPETING TASKS AS MEASURES OF INTELLIGENCE AND PREDICTORS OF JOB PERFORMANCE

1.1 Structure

This dissertation comprises five chapters. This chapter briefly sets the scene for the next four chapters. Chapters two, three and four cover the four studies undertaken for this dissertation. Each chapter is devoted to a separate study with the exception of Study 1, which comprises two parts; Part A (the pilot study) and Part B (the first study to include job performance data). Study 1 is covered in Chapter 2, Study 2 is covered in Chapter 3 and Study 3 is covered in Chapter 4. Chapter 5 forms the general discussion and conclusion.

1.2 Statement of Objectives

The competing task paradigm refers to tasks that require simultaneous performance of two cognitive activities (Stankov, 2003), such as remembering the number and location of balls in cups while, at the same time, the order of presentation of words. One task may be presented in one modality (e.g., a visual presentation of balls moving between cups) and one in another (e.g., spoken words), or both may be presented in the same modality; the key requirement is that mental operations need to be carried out within the same period of time.

Competing task paradigms have a long history in psychology research. They have been used in different domains of psychology, from cognitive to organisational, neuropsychology and studies of individual differences; in both experimental and human factors settings to test theories of attention and to explore aspects of applied cognition (Kahneman, 1973; Sauer, Wastell, & Hockey, 1999; Shirey & Reynolds, 1988; Spelke, Hirst, & Neisser, 1976). Their use has been widespread and varied but also sporadic, largely because they have traditionally been difficult and costly to construct and administer (Moreno & Mayer, 1999; Schmitt & Mills, 2001). For instance, to conduct research in this area in the 1970s expert assistance from technicians in acoustic laboratories was required (Stankov & Horn, 1980) or it was confined to settings equipped with sophisticated technology such as flight training centres (e.g., North & Gopher, 1976).

These difficulties are easing as the technology required for competing task construction becomes available to people working outside well-equipped laboratories (Moreno & Mayer, 1999; Wickens & Seidler, 1997). Increasingly, the value of this paradigm in applied settings is being realised, and in recent times competing tasks have been used in research related to job selection (Stankov, Fogarty & Watt, 1989).

A question of particular interest to this research study was the potential of competing tasks to predict job performance. To explore this question, it was necessary to explain how the abilities measured by competing tasks related to other well-established predictors of job performance, notably general mental ability and personality. The pilot study (Study 1 Part A) established the general test properties of the competing task measure, Study 1 (Part B) explored the relationship between competing task performance and other ability measures. Study 2 expanded the research to include personality measures and Study 3 investigated cross-cultural differences. The literature review that follows will further explore the competing task paradigm, outline the theory of intelligence that served as the framework for this research study, and explain how competing tasks relate to the various components of the this theory.

1.3 Broad Study Aims

The broad aims for the three studies are as follows:

1. To determine whether the Multi-Tasks test is a reliable measure.
2. To investigate the underlying factor structure of the Multi-Tasks test.
3. To determine the place of Multi-Tasks within the model of intelligence, with reference to The CHC Theory of Cognitive Abilities.
4. To determine the predictive validity of Multi-Tasks with reference to job performance.

5. To compare the predictive validity of Multi-Tasks with other well-known measures of job performance, including cognitive ability tests and personality measures.
6. To investigate the cross-cultural validity of the Multi-Tasks tests within the Chinese workforce.
7. To investigate group differences in performance on Multi-Tasks, including gender, age, level of education and organisational level.

CHAPTER 2

STUDY 1

PILOT STUDY (PART A)

AND

COMPETING TASKS AS MEASURES OF INTELLIGENCE AND PREDICTORS OF JOB PERFORMANCE (PART B)

Competing Tasks, the Multi-Tasks Test, Intelligence and Job Performance

2.1. Definition of Competing Tasks

Competing tasks are tasks that are performed simultaneously or with a short delay between them. It should be noted that 'competing-task' or similar terms such as 'dual-task' are sometimes used in methodologies where one task is a distractor or is secondary to the other (e.g., listening to music whilst working) or requires a motor rather than a cognitive response (e.g., pushing a button) (Schumacher & Schwarb, 2009). These are not the paradigms of interest in the current dissertation. The paradigms that are of interest involve demanding tasks which compete for attention, thus producing an information 'overload' (Stankov, 1983b). Good examples of such tasks are those from the aviation industry which has a long history of using competing task paradigms in their personnel selection processes. The industry requires in their pilots and air traffic controllers an ability to attend and respond to numerous pieces of information simultaneously, thus they have incorporated these tasks into psychometric testing for job candidates. Most recently, the US Federal Aviation Administration Air Traffic Controller Selection and Training battery (AT-SAT) includes two multi-tasking tasks, one which is directly relevant to the job role and involves an air traffic control scenario whereby candidates are required to keep track of numerous flights and one which is not directly job related but involves categorising letters and prioritising tasks according to a set of rules (Dattel & King, 2010; King, Manning, and Drechsler, 2006). Detailed explanations of these tasks are not provided, presumably so as not to give potential job applicants the opportunity to practice these tasks in advance.

The other area where competing tasks have been long been employed is in university-based research where such tasks have been developed and trialled on university students with a view to these tasks being adopted in personnel selection outside the university. In such research the competing task paradigms typically employ tasks such as memory tasks (e.g., recalling the order of presentation of a target word/letter or the number of times a certain word or letter string occurs), mental calculations (e.g., adding numbers, mentally swapping the order of numbers or letters) and auditory monitoring (e.g., recalling the number of times a tone of a certain pitch is presented) (Stankov, 2003; Wang, Proctor, & Pick, 2007). Participants are initially asked to do a single task, and then later they are given items from two of the tasks at the same time, either presented through different modalities, or both through the same modality with a short delay between presentation of the first and second task (Stankov, 2003). Participants are instructed to attend to both tasks equally. Task parameters such as speed of presentation and number of items in each task are manipulated in order to increase difficulty. Performance on the competing task paradigm is then compared to performance on the single tasks.

The mechanisms underlying performance on such tasks will be discussed in the next section.

2.1.1 Mechanisms underlying performance on competing tasks.

2.1.1.1 Capacity models of attention.

A significant amount of research has been devoted to the underlying mechanisms of competing task performance and numerous theories have emerged to explain this type of performance. Some theorists argue that two tasks can only be performed simultaneously by rapidly switching attention between them, while others argue that these tasks can be performed by dividing attention (see Fogarty & Stankov, 1982). Capacity models of attention discuss competing tasks in relation to the distribution of available limited resources. Theorists differ in terms of their focus in relation to the implications of a limited capacity system. Kahneman (1973) and later Hockey (1997), building on Kahmeman’s model, discussed the functional implications of limited capacity such as the way task demands are managed in order to accomplish goals. Hockey (1997) hypothesised that trade-offs among individuals’ goals, motivation, strategies and allocation of mental effort determine whether they will use adaptive or maladaptive strategies in the face of competing demands. For example, when faced with a need to perform two tasks simultaneously, individuals can intensify their mental effort; however the trade-off becomes increased anxiety and fatigue. Alternatively, they may lower their goals or adopt less effortful strategies, but at the cost of impaired performance.

These theories also propose that the way tasks are processed, whether they can be processed in parallel, or whether one task assists or interferes with the other depends on features of the tasks such as task similarity, temporal overlap or length of time between task presentation and perceptual overlap (Ellenbogen & Meiran, 2011; Miller, Ulrich, & Rolke, 2009).

Bottleneck theories are also limited capacity models but they focus on the structural implications of limited attentional resources. Bottleneck theorists propose that processing resources are scarce and can only deal with one task at a time, therefore information must be processed serially, rather than in parallel (Ruthruff, Van Selst, Johnston, & Remington, 2006). In their study using a competing task paradigm, Ruthruff et al.found that most participants responded to tasks in a serial manner. However, interestingly, four participants demonstrated the ability to bypass the bottleneck and respond to both tasks simultaneously.

The assumptions of capacity models of attention have been refuted by Stankov (1989), who provided examples in the competing tasks literature of findings that did not support them. This included research showing that performance on competing tasks can improve with practice, thus questioning the concept of a bottleneck and limited processing resources (Stankov).

While the theoretical mechanism underlying competing tasks may not be clear, it is apparent from any theoretical standpoint that task characteristics are important determinants of performance, as some tasks will be easier to undertake simultaneously than others. For instance, performance is likely to be better if one or both tasks can be processed quickly and autonomously, the delay between tasks or the modality of presentation allows for rapid attention switching, or if tasks do not compete for the same processing structures (Ellenbogen & Meiran, 2011; Fogarty & Stankov, 1982; Stankov, 1988). Tasks that place a higher burden on working memory may also be more likely to cause interference in multi-task processing than those that do not (Ketelsen & Welsh, 2010). This is unsurprising given that working memory has been shown to be strongly related to g (Kyllonen & Christal, 1990). This proposition will be further explored in the sections that follow. Strategy use can also be affected by task instructions, meaning that whether a person attends equally to each task or devotes most resources to one task at the expense of the other, or prioritises performing tasks quickly rather than accurately, for instance, may depend on what they are asked to do (Janssen & Brumby, 2010).

2.1.1.2 Individual differences.

There is also evidence of individual differences in the ability to use processing resources to undertake competing tasks. It has been argued that people differ in the amount of processing resources they possess, and that these differences in processing resources may account for individual differences in intelligence (Ben-Shakhar & Sheffer, 2001). That is, people with more overall processing resources may be more capable of performing competing tasks, as they have more resources to dedicate to the processing of each task (Stankov, 2003). Further, some individuals may utilise available resources more efficiently and therefore perform better than others, and some will find certain combinations of tasks easier than others (Fogarty & Stankov, 1982; Watson & Strayer, 2010).

2.1.1.3 Timesharing.

It has been hypothesised that a ‘time-sharing’ factor may explain individual differences in the ability to perform tasks simultaneously (Sverko, Jemeic, & Kulenovic, 1981 as cited in Fogarty & Stankov, 1982). Time-sharing is a hypothetical factor for which there is currently no conclusive proof. For this reason it is difficult to define. It has been referred to as a factor associated exclusively with performance on competing tasks, which is not apparent when the same tasks are undertaken independently and which is not an artifact of task instructions or presentation (Fogarty, 1987).

Fogarty and Stankov (1982) found some tentative support for a time-sharing factor in a study where they compared performance on individual psychometric tests with performance on a subset of these tests presented in a competing task format. Following the competing task presentation participants were asked to respond to one task initially (which was cued after its presentation to avoid participants focusing on what they considered to be the easier task), and to only provide the answer for the second task if they were sure of their response to the first. Performance on single tasks and the cued task in the competing task paradigm (primary scores) were highly correlated, indicating that they were measuring the same underlying construct, but another factor emerged that was unique to the secondary scores in the competing task paradigm. It appeared to represent a 'time-sharing' factor present only in situations where attention was divided between tasks. They cautioned, however, that this factor could also represent something to do with instructions or strategy use.

Evidence for a time-sharing factor was also found by Salthouse and Miles (2002). They explored individual differences in the ability to perform two tasks simultaneously across age groups. In their study participants performed three demanding 'primary' tasks independently and then each in combination with a visual tracking task. The primary tasks were: (a) keeping track of directions which involved verbal instructions as to current direction then a number of shifts after which the participant was to indicate the final direction; (b) competing number series which involved auditory presentation of a series of five numbers after which the participant had to indicate what number would complete the series; and (c) word pairs during which participants were presented with a set of word pairs and were asked to recall the second member of the pair when presented with the first. The competing task component was kept to a single task in an effort to maximise the opportunity to reveal a time-sharing factor by removing variability caused by having a range of measures. The visual tracking task involved manipulating a trackball in an attempt to keep a cursor on a randomly moving target. Participants were instructed to focus on performing the primary task but also to perform the visual tracking task to the best of their ability given that constraint. They found that measures of performance decrement in the competing task paradigms were correlated strongly with each other but relatively weakly with other cognitive measures. They concluded that the performance decrement represented a distinct time-sharing ability.

Expanding on the work of Fogarty and Stankov (1982), Fogarty (1987) investigated the existence of this factor using a similar methodology and a large selection of task combinations covering a range of broad abilities from different levels of the CHC Theory of Cognitive Abilities. A factor was extracted that encompassed a reasonable number of competing task measures, but almost half the competing task measures did not load onto this factor, indicating that the factor did not capture something unique to the competing task paradigm (Fogarty). Support for a time-sharing factor was therefore equivocal in this study. Research on time-sharing has been hampered by a number of issues, including the vague and somewhat contradictory use of this term in the literature. For instance, time-sharing has been assumed to exist in the context of competing tasks (and used interchangeably with terms such as multi-tasking and dual-task performance) rather than being empirically tested (e.g., Anderson, Taatgen, & Byrne, 2005; Watson & Strayer, 2010), or it has been defined as the ability to switch attention (rather than divide attention as per Stankov and Fogarty’s work) between tasks (Brown, 1998), or defined broadly as a set of skills unique to the competing tasks paradigm, including attention switching, efficient response strategies and parallel processing (Wickens, 1992). Studies have been criticised for using inappropriate methodologies such as exploratory factor analysis which may be inadequate to detect a time-sharing factor (Fogarty, 1987). In addition to methodological and theoretical issues, there are other difficulties that make it a hard factor to extract and measure, including individual differences in ways people approach the same competing tasks (Fogarty).

Individual differences in performance on competing tasks are therefore an important area of research. This will be explored within the context of the CHC Theory of Cognitive Abilities, a prominent theory which provides the theoretical framework for this dissertation. The following sections will introduce the CHC Theory and explain its relevance to competing tasks.

2.2. The CHC Theory of Cognitive Abilities

In the second half of the 20th Century, a theory of intelligence emerged that has been given a prominent position amongst intelligence theories. This is the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (McGrew, 2004), which is an amalgamation of Cattell and Horn's Gf-Gc theory (Horn & Noll, 1997) and Carroll's three-stratum theory (Carroll, 1997). The CHC Theory has been described as the "consensus psychometric-based model for understanding the structure of human intelligence" (McGrew, 2009, p.1). It is the theoretical basis for most of the most widely used psychometric tests of intelligence (McGrew, 2009).

Extensive factor analysis by Cattell and Horn revealed the existence of ten broad abilities in the intelligence literature (Horn & Noll, 1997), including fluid and crystallized intelligence (Gf and Gc, respectively). Other factors uncovered include a broad visualisation function (Gv), broad auditory function (Ga), short-term acquisition and retrieval (memory) (SAR), tertiary or long-term storage and retrieval (TSR), broad speediness function (Gs), correct decision speed (CDS) and in the extended model, quantitative (mathematical) knowledge (Gq) and language-based (reading and writing) skills and knowledge (Horn &

Noll, 1997). These factors formed the basis of the Gf-Gc theory. When it was developed, the Gf-Gc theory was described as being the most empirically grounded available theory of intelligence (Carroll, 1993).

Carroll (1997) also employed factor analysis in order to analyse data sets from intelligence research, and found that the range of human abilities fit neatly into three levels, or strata. The first stratum comprises specific abilities that people possess, such as reasoning and comprehension, while the second stratum contains eight broad structures, similar to those found in the Gf-Gc theory, that describe the first stratum abilities. Stratum three simply contains a general factor of intelligence.

Due to their similarities, as well as the amount of research that supported the conclusions made by Horn and Cattell (Horn & Noll, 1997; McGrew, 2004), and Carroll (1997), these theories were combined to create a single comprehensive theory of human cognitive abilities, called the CHC Theory of Cognitive Abilities (McGrew, 2004). The CHC theory is a hierarchical framework consisting of three strata: general intelligence or g (stratum III), broad cognitive abilities (stratum II) and narrow cognitive abilities (stratum I).

The CHC broad abilities are similar to those factors in the Gf-Gc theory, with the following alterations: SAR has been replaced with the term Gsm (short-term memory), TSR has been replaced with the term Glr (long-term memory) and the speediness (correct decision speed) factor has been broken into two factors, Gs which refers to cognitive processing speed, and Gt which is decision and reaction speed (McGrew, 2009). Thus, the CHC theory of cognitive abilities consists of Gc, Gf, Ga, Gv, Gsm, Glr, Gs, Gt, Gq and Grw (McGrew, 2004). The model also includes six “tentatively identified broad ability domains” (McGrew, 2009). These were added following a comprehensive review of the factor analytic research on intelligence between 1993-2003 (McGrew, 2005), the results of which suggested that all sensory modalities should be included in a model of human cognitive abilities. These further domains are general domain specific knowledge (Gkn), tactile abilities (Gh), kinesthetic abilities (Gk), olfactory abilities (Go), psychomotor abilities (Gp) and psychomotor speed (Gps) (McGrew, 2009).

Figure 1.1 outlines the structure of the three theories, lists their components, and demonstrates how they relate to each other.

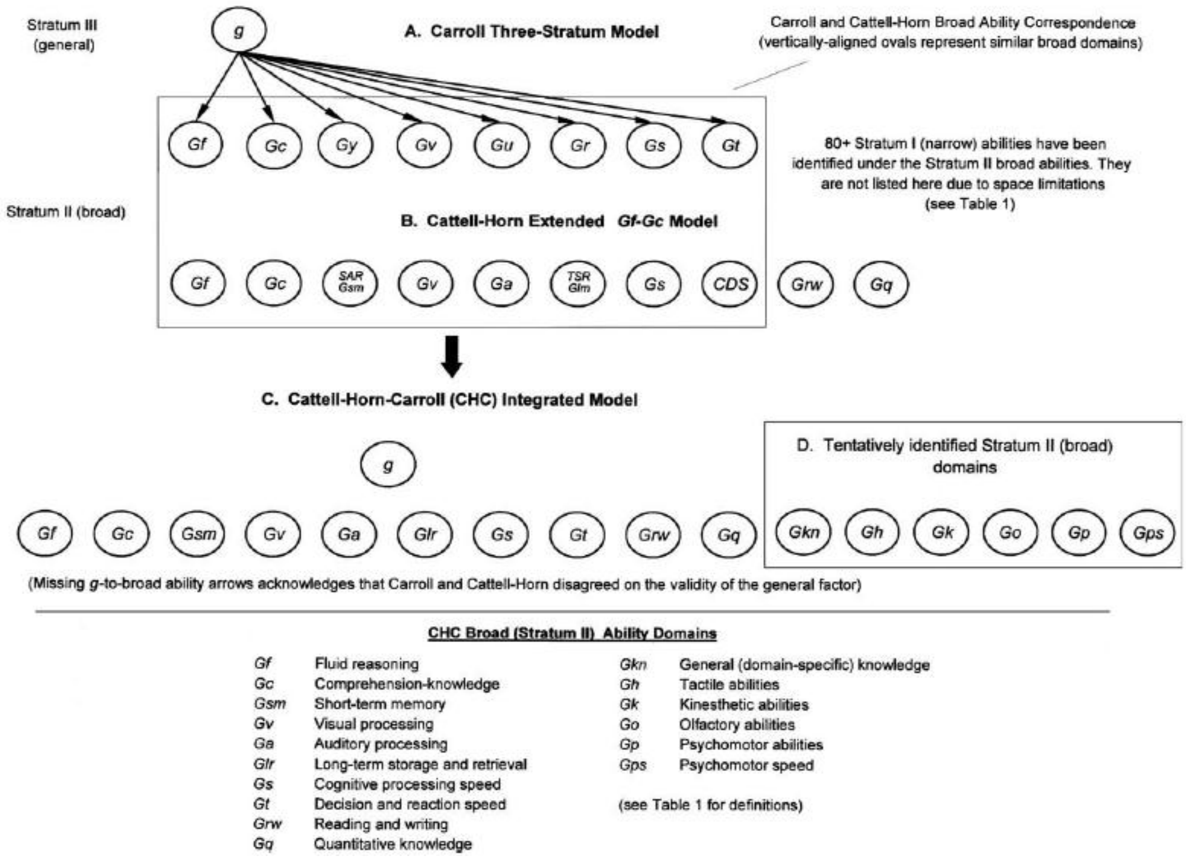


Figure 1.1. Schematic representation and comparisons of Carroll's Three-Stratum, Cattell–Horn's Extended Gf–Gc, and the integrated Cattell–Horn–Carroll models of human cognitive abilities.

Source: McGrew (2009).

2.2.1 Definitions of broad abilities.

All three theories (The CHC Theory, Cattell and Horn's Gf-Gc theory and Carroll's three-stratum theory) include Gf and Gc as key components of intelligence. Gc, or comprehension-knowledge in the CHC model (referred to as crystallized intelligence in Carroll's theory), refers to knowledge that is gained through formal education (Horn & Noll, 1997; Stankov, 1988). Individual differences in Gc reflect differences in its acquisition and storage. Gc is thus measured by tests that call upon learned information from long-term memory (Stankov, 1988).

Gf (named fluid reasoning in the CHC model and fluid intelligence in Carroll's theory) is an outcome of personal experience and incidental learning that is unaffected by education (Stankov, 1988), and is tested with tasks that involve working memory, attention, reasoning and deduction (Stankov, 2003). As described by Horn and Noll (1997), Gf refers to abilities that are measured in tasks requiring inductive, deductive, conjunctive, and disjunctive reasoning to arrive at understanding relations among stimuli, comprehend implications, and draw inferences (p. 69).

A third component of intelligence that is important in the current dissertation in relation to its role in competing tasks performance is Gsm. This component is simply the ability to apprehend and maintain awareness of a limited number of events that have occurred very recently (McGrew, 2009). It is limited in capacity and loses information quickly unless other cognitive resources are employed to maintain the information in immediate awareness (McGrew, 2009). Gsm is measured through tests of immediate memory span (Stankov, 1988). Gsm is the 'passive' component of working memory (WM) (Stankov, 1988). WM is composed of a temporary memory store (Gsm) which includes a mechanism for the cognitive rehearsal of information, and an executive mechanism which manages attention and active content (i.e., manipulating the information in the temporary memory store), particularly in the presence of interference (Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2002; Stankov, 1999). Refer to Table 1.1 for brief definitions of the other key broad ability areas.

Table 1.1
CHC broad ability domain terms, corresponding Cattell–Horn and Carroll broad ability terms, and definitions. Adapted from McGrew (2009).

CHC model	Cattell-Horn theory	Carroll theory	Definition
Visual processing (Gv)	Visual processing (Gv)	Broad visual perception	The ability to generate, store, retrieve, and transform visual images and sensations.
Auditory processing (Ga)	Auditory processing (Ga)	Broad auditory perception (Gu)	A wide range of abilities involved in the interpretation and organization of sounds, such as discriminating patterns in sounds and musical structure (often under background noise and/or distorting conditions) and the ability to analyse, manipulate, comprehend and synthesise sound elements, groups of sounds, or sound patterns.
Long-term storage and retrieval (Glr)	Long-term storage and retrieval (TSR, Glm)	Broad retrieval ability (Gr)	The ability to store and consolidate new information in long-term memory and later fluently retrieve the stored information (e.g., concepts, ideas, items, names) through association.
Processing speed (Gs)	Cognitive processing speed (Gs)	Broad cognitive speediness (Gs)	The ability to automatically and fluently perform relatively easy or over-learned elementary cognitive tasks, especially when high mental efficiency (i.e., attention and focused concentration) is required.
Reaction and decision Speed (Gt)	Correct decision speed (CDS)	Processing speed (RT)	The ability to make elementary decisions and/or responses (simple reaction time) or one of several elementary decisions and/or responses (complex reaction time) at the onset of simple stimuli.
Reading and writing (Grw)	Reading and writing (Grw)	Carroll included reading and writing narrow abilities under Gc	The breadth and depth of a person's acquired store of declarative and procedural reading and writing skills and knowledge.
Quantitative knowledge (Gq)	Quantitative knowledge (Gq)	Carroll included math achievement factors in a chapter on “Abilities in the domain of knowledge and achievement”.	The breadth and depth of a person’s acquired store of declarative and procedural quantitative or numerical knowledge

2.2.2 Working memory and the CHC Theory.

WM has received significant attention in the intelligence literature due to its relationship with processing concepts such as sustained attention, memory span, and processing speed (McGrew, 2004). The WM system must hold a substantial amount of information, often to be manipulated, which is necessary for solving complex problems (McGrew, 2004). Therefore is it strongly associated with Gf. Neuropsychological and intelligence research studies have reported correlations of 0.6 to 0.8 between WM and Gf (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, et al., 1999; Kane, Hambrick, & Conway, 2005; Kyllonen & Christal, 1990). The importance of WM is also demonstrated by its strong relationship with overall intelligence. In a series of three studies, Colom, Rebollo, Palacios, Juan-Espinosa, and Kyllonen (2004) found an average correlation of .96 between WM and g for 594 participants (psychology undergraduates and Air Force recruits).

There has been some debate about whether working memory (WM) should be included as a component of the CHC theory (McGrew, 2004). However, research has shown that while it is an important cognitive construct, WM is not comparable to the other narrow trait-like CHC components. It does not depend on factor analysis for its identification and it was not developed as an individual differences factor but a construct to explain experimental findings from memory studies (McGrew, 2004; 2009). McGrew (2009) discusses its importance in understanding new learning and performance of complex cognitive tasks. Further, it is likely to be an important construct in the context of competing tasks.

2.2.3 Mental speed- the key to understanding intelligence?

Mental speed or the speed at which one can perform basic mental operations as an indicator of intelligence has been prominent in the study of human cognitive abilities (Nettelbeck, 1994; Nyborg, 2003; Stankov & Roberts, 1997). Indeed, some researchers have suggested that mental speed may be the key process underlying intelligence (e.g. Kail, 1991). The claim arises from the information processing model of human intelligence, whereby cognition is constrained by limited processing resources, particularly in working memory, and that the speed with which information is processed is critical because it determines how quickly and efficiently resources can be allocated to cognitive tasks (Kail, 1991). Further, processing speed has been found to account for a large proportion of age-related decline in ability to perform complex mental operations (Hambrick, Oswald, Darowski, Rench & Brou, 2010).

Correlational findings amongst speed measures, Gf and Gc, do support this assertion about the importance of processing speed, however. Correlations between the speed component of the CHC theory of cognitive abilities (i.e., Gs), and Gc, have been found to be non-significant (Roberts, Beh, Spilsbury, & Stankov, 1991; Stankov & Roberts, 1997). If correlations between mental speed and Gc were non-significant, one would then expect speed to correlate with Gf measures. Stankov (2003) reports that this is not the case. Further, the authors found that mental speed alone, or indeed any other cognitive process (Hunt, 1980), could not explain correlations between cognitive tasks and intelligence (Stankov & Roberts, 1997). Correlations between components of cognitive ability, such as mental speed, and intelligence appear to reach a barrier of approximately 0.30, above which a single component cannot explain any more of the variance in intelligence (Stankov & Roberts, 1997). Lastly, both theories of intelligence which have been combined to produce the CHC theory of cognitive abilities, postulate the existence of eight or nine broad abilities (McGrew, 2004).

Processing speed is considered amongst these other abilities in both theories, rather than being singled out as the main process.

The relationship between intelligence as described above and the competing tasks paradigm will now be outlined in some detail. This outline will then lead into a discussion about how competing tasks can be useful in predicting job performance.

2.3. Competing Tasks and Intelligence

2.3.1 CHC theory and competing task performance.

One of the aims of previous research has been to examine in more detail the relationship between competing task performance and the various components of intelligence.

A number of studies have examined the relationship between competing tasks and various components of intelligence. Gf, processing speed and WM have received the most research attention, with the strongest associations being found between Gf, WM and competing task performance. The role of processing speed remains unclear, though worthy of discussion because of the attention this concept has received in both intelligence and competing tasks research.

2.3.2 Competing tasks, Gf and WM.

Research indicates that the Gf component of the CHC model plays the greatest role in predicting performance on competing tasks (e.g., König, Bühner, & Mürling, 2005; Stankov, 1988). WM has also been found to play an important role in competing task performance, largely resulting from its relationship with Gf.

König, et al., and Mürling (2005) investigated WM and Gf using the competing task paradigm. These authors found that WM, attention and Gf emerged as the most important predictors of performance on the competing task measure, with WM resulting in the highest correlations with performance. WM yielded the highest correlations with all components of the competing task measure and explained incremental variance in the hierarchical regression analysis that could not be explained by Gf. They surmised that because of the complexity of the competing task paradigm it was the higher level cognitive processes, specifically executive control of attention that accounted for most of the individual differences in performance. They concluded that WM tasks may be important in job selection for roles requiring a high amount of multi-tasking.

In a number of studies the higher level ‘central executive’ function (as opposed to the passive component) of WM has been tested by introducing interference into tasks; thereby assessing the ability of participants to maintain goals in the presence of this interference (Engle, 2002; Newman, Keller, & Just, 2007). Competing tasks are one such paradigm. Neuropsychological research has found an interesting set of relationships amongst WM, Gf and competing task performance. Specifically, studies have found that tasks testing WM and Gf result in similar activation of the prefrontal cortex (PFC), particularly the dorsolateral PFC (dPFC) (Kane & Engle, 2002, 2003), which is thought to perform as a central executive mechanism, allowing for the coordination and management of goals and task-relevant information in the presence of interference (Dempster & Corkill, 1999; Engle, 2002; Kane & Engle, 2002, 2003). Interestingly, performance on competing tasks has also been found to result in activation of the PFC, although the pattern of activation has been found to differ across studies (Adcock, Constable, Gore, & Goldman-Rakic, 2000; Bunge, Klingberg, Jacobsen, & Gabrieli, 2000; Collette et al., 2005; D'Esposito et al., 1995; Just et al., 2001; Klingberg, 1998; Newman, et al., 2007; Szameitat, Schubert, Müller, & von Cramen, 2002). A number of studies using functional magnetic resonance imaging (fMRI) have demonstrated that the brain adapts to competing task performance (where the tasks are of sufficient complexity to be unable to be processed automatically) by employing additional resources from the PFC. These resources were not activated when the tasks comprising the paradigm were presented individually and were thus unique to the competing task paradigm (Buchweitz et al., in press; Jaeggi et al., 2003; Kondo, Osaka, & Osaka, 2004)

The correlation between Gf and WM is not always supported by findings from intelligence research. Stankov and Myors (1990) found that WM load was more strongly associated with short-term memory (or SAR, renamed Gsm in later models) and Gc than Gf. Increases in WM load were not found to lead to increased correlations with intelligence, in fact the more complex the task and the higher demands on working memory, the lower the correlation with IQ.

Conflicting research findings may be due to variations in the definition of WM used in individual differences versus neuropsychology research, as well as to differences in the focus of the research, the intelligence tests used, the nature of the competing tasks, and the way performance is measured. Stankov and Myors (1990), for instance, did not use the competing task paradigm in their study and their intelligence measure (WAIS-R) was not developed with the CHC theory as the key underlying theoretical model. The WAIS-R has also been claimed to be primarily a measure of Gc and not Gf (Kane & Engle, 2002). While both branches of psychology define WM in terms of two components – the passive short-term store and the active executive control mechanism – the focus of the neuropsychological research has been on the processes associated with the active component. Meanwhile, much of the individual differences research has focused on WM as a limited capacity system, related to theories of attention (Oberauer & Göthe, 2006).

There is reason to believe that there may be a relationship between WM and competing task performance but that further research is required, particularly tasks that involve the active component (executive control mechanism) of WM.

2.3.2.1 Competing tasks and processing speed.

Processing speed measures have been commonly used in studies of competing tasks, likely because of their relationship with WM and because they have received much research attention in explaining individual differences in intelligence. The theory is that individuals who can process information quickly possess a more efficient processing system and are therefore better able to deal with increases in complexity and/or demands on WM, as is the case with competing tasks measures. This suggests that when there is a large information flow, as is the case with competing tasks, faster speed of processing is likely to be beneficial.

However, in a review of the literature, Stankov and Roberts (1997) reject the notion of mental speed as the process that can explain competing task performance. Mental speed on its own has been unable to explain correlations of competing tasks with intelligence (Roberts, et al., 1991), and is often studied in conjunction with other confounding factors such as complexity. Mental speed itself is often hard to isolate from other types of speed, such as decision speed or reaction time (Stankov & Roberts, 1997).

While research suggests a positive relationship between Gf and performance on competing tasks, it is important to consider why such a link might exist, and whether, if this is the case, a competing tasks measure has any advantage over traditional Gf measures. The following sections explain the relationship between competing tasks and Gf in terms of the underlying cognitive processes, thereby demonstrating the importance of competing task measures over traditional ability measures in terms of their complexity and utilization of a range of cognitive abilities including processing resources and attention, skills which are required in the modern complex workplace.

2.3.3 Cognitive mechanisms underlying competing tasks.

2.3.3.1 Complexity.

It may be that added complexity of competing tasks compared with single tasks explains the higher correlation of competing tasks with intelligence, particularly fluid intelligence.

It is important to distinguish between complexity and difficulty, terms which are often used interchangeably in the competing task literature but which refer to distinctly different concepts (Spilsbury, Stankov, & Roberts, 1990). According to Stankov (2003), difficulty is conceptualised as the percentage of people who can accurately answer an item, while complexity involves any manipulation that increases the WM load and correlation of the task with measures of intelligence.

To illustrate what is meant by complexity manipulation, the Swaps test is a psychometric test where participants are presented with three letters and are instructed to mentally swap two of the letters around (Stankov, 2000, 2003). This task can be used as a measure of complexity by adding up to three more swaps that the participant must mentally perform before answering. This test is thought to be a good measure of fluid intelligence (Stankov, 2003), as Stankov (2000) found a uniform increase between Gf and the complexity of the task.

Studies with the Swaps test have found that performance decreases as the number of swaps increases (see Stankov, 2003). However, correlations with measures of fluid intelligence have been found to increase as more swaps are added to the task. Furthermore, studies with this test have revealed a fanning-out effect in the data, such that, as the number of swaps increased, the difference between individuals with high- and low-Gf also increased (Stankov, 1999, 2003). These findings indicate that complexity and Gf are strongly related.

Further evidence to support the relationship between complexity and intelligence comes from information processing research. This domain differs from other competing task research in that it considers complexity in terms of ‘bits’ of information (Roberts, et al., 1991). There appears to be a linear relationship between intelligence and the number of bits of information that are able to be processed (Roberts, et al., 1991). Importantly, Roberts et al. found that adding another task so that two tasks were presented in a competing manner adds one extra bit of complexity. This finding was supported by Spilsbury et al. (1990), who also found that competing tasks were manipulations of complexity, while other tasks often used in the literature to test intelligence appear to be more related to difficulty (e.g., making the text smaller on a perceptual task makes it more difficult, but not more complex) (Spilsbury et al., 1990). To help differentiate, there appears to be a quantitative change in a task when difficulty is increased, and a qualitative change when complexity is manipulated (Ben-Shakhar & Sheffer, 2001; Roberts, et al., 1991).

Research by Roberts et al. (1988) found that competing tasks resulted in higher correlations with Gf than the two components of the competing task presented separately. Furthermore, correlations between Gf and competing tasks increased as extra bits of information were added. Roberts et al. (1991) replicated these findings, as increasing complexity related to an increase in correlations of the task with Gf measures. However, correlations of performance with Gc while showing a positive trend, were mostly non-significant, and when Gf was partialled out all approached zero. SAR (short-term acquisition and retrieval, renamed Gsm in later models) was not strongly correlated with performance and correlations reduced under the competing task paradigm compared with the single task presentation (Roberts et al., 1991).

It would appear, therefore, that complexity manipulations have the strongest impact on measures of fluid intelligence, such that increasing the complexity of a task (e.g., by increasing the number of components involved) increases the correlation with Gf. It has been suggested that WM may play a mediating role in the relationship between complexity and Gf. Spilsbury et al. (1990) suggest that increasing the complexity of a task increases its WM load, which then strengthens the relationship with Gf. Evidence to support this relationship has been found in the intelligence literature. For example, Carpenter, Just and Shell (1990) found that commonly used tests of Gf, the Raven’s Standard Progressive Matrices and the Tower of Hanoi, measure abilities which are thought to be associated with WM load, such as executive control of attention and the generation and maintenance of goals, particularly as the tests progress and becomes more complex. Their analysis of those individuals who were able to perform both tasks well, compared to those who performed poorly, lead the authors to conclude that successful completion of the more complex items requires employing these higher level processes (executive control and goal maintenance), which then results in better performance on tasks of Gf (Carpenter et al., 1990).

2.3.3.2 Pools of resources.

Another theory that has been put forward to explain performance in competing tasks is the multiple pools of resources theory (see Kramer, Wickens, & Donchin, 1985). This theory comes from the finding that competing tasks do not all interfere with each other to cause performance decrements (Stankov, 1983), and that some task combinations actually enhance performance. This pools of resources theory posited that there were many different clusters of resources which can be utilised when processing multiple tasks at the same time.

Evidence for the existence of multiple pools of resources comes from multimedia research, in which audio visually presented information was found to result in greater knowledge acquisition than information presented in either an auditory or visual manner (Mayer & Moreno, 1998; Moreno & Mayer, 1999). This is explained in the literature in terms of a multimedia dual-processing theory, which posited the existence of two independent processing systems for visual and auditory information that can be used concurrently to process information (Brünken, Steinbacher, Plass, & Leutner, 2002; Mayer & Moreno, 1998). This was further supported by the finding of decreases in performance when information is presented in a visual-visual manner (Brünken, et al., 2002; Mousavi, Low, & Sweller, 1995).

Research relating to competing tasks and intelligence has also reported an effect of task modality. Fogarty and Stankov (1988), for example, found that audiovisual competing tasks loaded higher on intelligence than did auditory-only competing tasks. Structural interference is likely to have introduced an additional difficulty element in the auditory-auditory split. These results have important ramifications for the optimal design of competing tasks tests as measures of intelligence.

2.3.3.3 Attention.

Many studies (e.g. Stankov et al., 1989; Fogarty & Stankov, 1982; 1988; Stankov, 1983b) discuss the importance of divided attention in relation to competing task performance. The ability to divide one's attention is also necessary for many tasks assessing Gf. König, et al., (2005) found that measures of simple attention (reaction time) and divided attention predicted two of the three components of their competing task measure, which were answering questions and solving verbal, figural and numerical problems. Attention did not predict the speed component of the measure. In explaining their results, the authors concluded that it is the ability to control attention, rather than attention per se that is important in a competing task paradigm.

In the literature on attention, competing tasks have often been used to examine the relationship between intelligence and/or performance with age. For instance, Lorsbach and Simpson (1988) found an age-related decrement in performance on competing tasks for older adults, although these results appear to be confounded by complexity manipulations. The authors suggested that these results may reflect attention differences between younger and older participants (Lorsbach & Simpson, 1988).

Results from a study into the effects of competing tasks on performance in children add to the findings regarding the possible importance of attention. In a study comparing older and younger children’s abilities to split attention, Irwin-Chase and Burns (2000) found that, when performance on single tasks was controlled for, both age groups showed the ability to divide attention across the tasks when tasks were given equal priority. However, when one task was given more priority (such that attention had to be divided 75% / 25% across both tasks), older children were better able to allocate attention accordingly. Irwin-Chase and Burns concluded that the differences in performance were a consequence of younger children being less able to manage attention.

The concept of a specific ability to divide and manage attention has also been considered in relation to age in the intelligence literature. That is, the decline in the Gf component of intelligence over time has been suggested to relate to the capacity for dividing and maintaining attention (Horn & Noll, 1997). Thus attention, and specifically the ability to focus and divide attention, has been found in the literature on competing tasks to relate to both intelligence and enhanced performance.

Having defined competing tasks, discussed the paradigm within the theoretical framework of the CHC Theory of Cognitive Abilities and outlined a number of constructs that are important to consider in relation to competing task performance, namely Gf, WM, complexity, attention and processing speed, the discussion now turns to the predictive validity of competing tasks in job settings. This section will commence with a review of the literature discussing what is known about the relationship between intelligence and job performance.

2.4. Intelligence and Job Performance

Psychometric tests of intelligence have been shown to be the best single predictors of job performance (Bertua, Anderson, & Salgado, 2005; Gottfredson, 1986, 1997; Hunter, Schmidt, & Judiesch, 1990; O'Reilly & Chatman, 1994; Salgado et al., 2003; Schmidt & Hunter, 1992; Schmitt & Mills, 2001). In a review of the intelligence and work literature, Gottfredson (1997) found that tests of general intelligence (i.e., Gf and Gc) were better predictors of performance compared to other predictors, especially in complex work situations.

General intelligence has been found in both cross sectional and longitudinal studies to be strongly associated with occupational level (Jensen, 1998; Judge, Higgins, Thoresen & Barrick, 1999) and income (Judge et al., 1999; Murray, 1998). Cross-sectional studies show mean correlations between general mental ability (GMA) and occupational level as high as .95 (Jensen, 1998). GMA has also been found to consistently predict later income, even when variables such as socioeconomic status, quality of schooling and family background are well controlled (Murray, 1998). In relation to specific performance measures, Hunter (1980, as cited in Hunter and Hunter 1984) and Hunter and Hunter (1984) reviewed over 400 validity studies involving over 32,000 employees across the occupational spectrum and found that correlations between GMA and 'on the job' performance ranged from .40 to .58 for all except the lowest level jobs (for which the correlation was .23).

In their review, Schmidt and Hunter (2004) demonstrated that the relationship between GMA and work performance was mediated by job knowledge. In other words, they showed that people higher on GMA were better able to acquire job knowledge which resulted in better performance. There was a direct relationship between job performance and GMA (.31 for civilian jobs and .15 for military jobs), but this was much smaller than the relationship between both GMA and job knowledge (.80 and .63 respectively) and job knowledge and job performance (.56 and .61).

Success in managerial roles may have a curvilinear trend with intelligence. Ghiselli (1963), for example, found that increases in intelligence related to increases in managerial success but reached a ceiling point where, as intelligence increased, success decreased. A similar finding was reported by Gill (1982), in relation to intelligence and managerial skills, such as decision making and prioritising. Further, while researchers such as Salgado et al. (2003) and Hunter and Hunter (1984) found that a positive relationship between GMA and managerial performance, other researchers such as Colonia-Willner (1998) found that scores on measures of Gf did not significantly predict managerial ability, while experience and workplace-specific knowledge did.

Job research has not always kept up with intelligence research, thus creating a gap between theory and application. A common shortcoming is adoption of a narrow or unitary concept of intelligence (i.e., too few measures or tasks that measure only one component, such as Gc). Accounting for the multidimensional nature of cognitive abilities in job selection is likely to yield more accurate operational validities for selection tests and help researchers better understand how cognitive abilities relate to performance at work. This is particularly relevant as job roles are becoming less defined and more unstable, and this trend is likely to continue into the future (Bertua, et al., 2005). It is possible that components of intelligence, such as fluid intelligence, may emerge as more valid predictors of job performance than overall intelligence.

Overall, the bulk of the research tends to show that as tasks become more complex there is a greater involvement of GMA. Other factors such as job knowledge and novelty may also be important, particularly in managerial jobs.

2.5. Competing Tasks and Job Performance

Competing tasks may prove invaluable to the assessment of intelligence for personnel selection reasons due to the superior correlations of competing tasks over single tasks with multidimensional models of intelligence (Ben-Shakhar & Sheffer, 2001; Fogarty & Stankov, 1982; Roberts, et al., 1988; Stankov, 1988, 1989, 2003). In a range of studies, competing task performance has been shown to be strongly correlated with a general intelligence factor, and to have a stronger relationship with the higher order factor than individual component tests (Ben-Shakhar & Sheffer, 2001; Fogarty & Stankov, 1982; Stankov, 1988). This has been demonstrated, for instance using the Wechsler Adult Intelligence Scale – Revised (Stankov, 1988), a measure of Gf (Fogarty & Stankov, 1982) and a measure comprising verbal reasoning, quantitative reasoning and English skills (Ben-Shakhar & Sheffer, 2001).

Competing tasks have also been demonstrated to be better able than single tests to discriminate between people in low- and high-ability groups (North & Gopher, 1976), and within high-ability groups themselves (Fogarty & Stankov, 1995). The latter is particularly important because cognitive tests often do not discriminate well between those at the higher end of the intelligence/ability scale because of ‘ceiling effects’ (Fogarty & Stankov, 1995), i.e., when a number of individuals obtain the maximum (or close to maximum) score on a task, results do not reveal differences between the ability of these individuals. Such an outcome makes it difficult, for example, to differentiate between an individual with the potential to be a good manager and one who may be an excellent manager.

A theoretical basis for differences in performance on competing tasks between ability groups has been provided by Stankov (1988, 1989) who advised that high ability groups learn more rapidly to choose efficient cognitive strategies to deal with added complexity of the competing task. Other theories have focused on attentional resources and hypothesised that high ability individuals are better able to allocate their attention, or have better attentional resources (Ben-Shakhar & Sheffer, 2001).

Research concerning competition between tasks in the workplace indicates that many individuals are placed in situations where they have to cope with a significant degree of complexity and cognitive demand in the workplace. Complexity is a particularly important consideration given the relationship between task complexity and intelligence, particularly Gf. Excessive cognitive demand, including the requirement to concentrate on various sources of information and coordinate multiple tasks is common to many jobs in contemporary society. Working on two deadlines at once, monitoring information provided by colleagues in various time zones, talking on the telephone whilst searching for information on the Internet are just a few examples of the demands of the modern workplace. A competing task paradigm uniquely captures the skills required to cope with such complexity (Xie & Salvendy, 2000). The importance of competing tasks has been demonstrated in various complex work situations, including pilots (Damos & Smist, 1982; North & Gopher, 1976), astronauts (Sauer, Wastell, & Hockey, 1999) and physicians (Chisholm, Dornfeld, Nelson, & Cordell, 2001).

For example, in a study of physicians, Chisholm et al. (2001) found that emergency and primary care physicians experienced frequent interruptions which disrupted thought flow and caused errors such as incorrect medication dispensing. Similarly, the increasing complexity of flight missions and equipment that pilots must deal with indicates the importance of these individuals being able to cope with several, often conflicting, tasks (Damos & Smist, 1982; North & Gopher, 1976), which is not adequately tested by research utilising only single-task methods (Sauer et al., 1999). Carretta (1987) found that a test using the competing task paradigm (tracking and cancelling digits) given to US Air Force Officers to be a highly reliable and useful measure. Participants who performed well on the task were more likely to be recommended for more advanced flight training.

In Fogarty and Stankov (1982) and Stankov et al. (1989)'s research it was found that single tasks did not significantly predict the performance of employees in an Australian consultancy firm (as measured by performance ratings completed by high level officers), however, when these same tasks were combined into a competing task paradigm, the predictive power of these tasks became significant (Stankov et al., 1989).

Hambrick et al. (2010; Hambrick, et al., 2011) devised a competing task paradigm that aimed to emulate the multi-tasking demands of the modern workplace (a 'synthetic' work task). The tasks were arithmetic (simple addition), memory search (identifying whether a letter that appears on screen was one of a list of letters presented earlier), auditory monitoring (responding to a tone of a certain pitch and ignoring a tone of a different pitch) and visual monitoring (monitoring a needle on a petrol gauge meter and resetting it before it reaches empty). The tasks were presented with a short inter-stimulus interval between them and the aim was to coordinate performance of all the tasks to maximise their score. Different tasks were identified as priorities on different trials to increase the ecological validity of the test. Outcomes were compared to scores on measures of WM and processing speed.

The authors found a strong relationship between performance on the synthetic work task and WM and a weaker relationship with processing speed. Further, they advised that there was a significant effect of practice, as indicated by higher scores for those who had more extensive video game experience (Hambrick et al., 2010). Whether this indeed represented a practice effect is debatable, but it did suggest that there is a general, trainable effect on multi-tasking over and above cognitive ability that appears to relate to experience with similar tasks. As was discussed in the paper, the main limitation of the research was that performance on the task was not validated against real life work performance (Hambrick et al., 2010). In a later article Hambrick et al. (2011) compared performance on the synthetic work task to scores on the Armed Services Vocational Aptitude (ASVA) Battery, which has been found to predict performance in military jobs. WM played an important role in predicting performance. Memory updating, a component of WM, partially accounted for the relationship and added incrementally to the prediction of multi-tasking performance, beyond overall ASVAB scores alone.

In addition, competing task presentations maybe more useful than single tasks in predicting managerial potential, a theory that was supported by Stankov et al. (1989). The skills uniquely measured by the competing task paradigm, including ability to selectively attend to important information and divide attention between tasks is a key requirement of managerial roles (Stankov et al., 1989). While researchers differ as to the proposed mechanisms underlying competing task performance, such as efficient strategy use or attentional allocation, any or all of these skills are likely to be important in managerial roles.

Increasingly, research indicates that the competing task paradigm show promise as useful predictors of job performance, both due to their high correlation with mental ability and with measures used in job selection research and because the underlying abilities they exploit have been shown to be related to better job performance, particularly in complex roles. Available research suggests that fluid intelligence and working memory are likely to be the most important underlying components to consider. Processing speed does not appear to play an important role in predicting job performance.

When considering the use of competing tasks to measure intelligence and/or predict real-world performance, it is important to consider a number of methodological issues, including the properties of tasks used and the effects of practice.

2.5.1 Methodological considerations.

There are a number of methodological issues to do with task construction and the effects of practice and feedback or incentives to consider when designing competing tasks, interpreting performance and considering implications for the workplace. The advantages and disadvantages of the Internet based format employed in this study will be outlined.

2.5.1.1 Task properties.

Task similarity, modality of presentation and stimulus difficulty, novelty and engagement are important considerations, which appear to be particularly important in the competing task paradigm. In relation to task properties, it has been found that task similarity may adversely affect correlations of tasks with measure of intelligence. For instance, Brünken et al. (2002) found that visual-only presentation of stimuli resulted in higher cognitive load and lower knowledge acquisition, while audiovisual presentation of the same information resulted in better performance. These findings were supported by many other studies (Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, et al., 1995). Stankov (2003) also found that selecting tasks which involve different abilities, such as spatial and verbal tasks, may be presented together to increase the complexity of the task without overwhelming cognition.

However, it should be noted that there is evidence suggesting that different types of visual stimuli, such as pictorial or verbal, can be processed independently. In a study of event related potentials (ERPs) associated with pictorial and verbal stimuli, Greenham, Stelmack and Campbell (2000) found that ERP waves for word processing described a distinctly different pattern of neural activation compared to those for image processing. This finding suggests that words and pictures may be processed independently (Greenham, et al., 2000), and therefore that visual-only presentation of multiple stimuli may not necessarily lead to performance decrements.

The discrepant findings of Brünken et al. (2002) and Greenham et al. (2000) may be explained by a spatial contiguity effect, wherein learning increases when information is presented in the same visual space, rather than being separated (Moreno & Mayer, 1999). This finding indicates that information may be able to be presented during competing tasks in a visual-only manner, but that it is important to combine the two elements within the same visual space so that excessive scanning is unnecessary.

Interestingly, a relationship between novelty and intelligence has also been found. Sternberg (1982) employed a novel concept-selection task in order to assess reasoning. Correlations between the task and intelligence were found to be higher than correlations of intelligence and traditional tests of cognitive ability (Sternberg, 1982). Therefore the novelty of the competing task paradigm may be an important contributor to its relationship with intelligence, especially Gf. The literature into the effect of engagement on performance revealed that recall of information by adults was increased when the material to be learned was engaging. Furthermore, interest has been found to result in a decrease in the amount of resources required for processing information (Shirey & Reynolds, 1988). Task difficulty has also been found to be an important consideration, as when faced with competing demands, individuals tend to allocate more resources and effort to the task they perceive to be more difficult (Northcraft et al., 2011).

Combining visual and auditory presentation of material in a competing task paradigm can present a number of challenges. These include the need for equipment such as headphones, addressing concerns about sharing headphones particularly in a work context where there are occupational health and safety regulations and technical issues associated with ensuring that the presentation of the visual and sound stimuli is well coordinated.

2.5.1.2 Practice.

Competing task studies have found a general trend for practice to lead to increased performance over time. For example, in an experiment where two tasks were given unequal weighting, such that one task was given priority over the other, Ruthruff, Van Selst, Johnston and Remington (2006) found that performance was better on competing tasks over single tasks when either of the single tasks, or the competing task itself, were practised. Stankov (1989) reviewed research suggesting that competing tasks performance shows a positive effect of practice after a certain number of trials. That is, the difference in performance between single and competing tasks at the beginning of practice trials was found to be greater than the difference after eight trials, by which time performance differences due to competition were no longer significant (Stankov, 1989). This research is encouraging, as it implies that, not only can people improve their performance on competing tasks, but that the ability to succeed at competing tasks may be trainable. This research also added to the debate regarding underlying theories of competing task performance and its relationship to intelligence. In particular, it challenges the bottleneck processing theory, as if practice can improve performance then it is less likely that processing is constrained by a ‘bottleneck’.

The outcomes of this research also implies that people who must take on roles in the workplace where they have to perform multiple tasks together may excel if they first undertake a training program to learn how to cope with competing tasks.

2.5.1.3 Feedback and incentives.

It has been shown that feedback and other organisational practices can influence the way in which individuals allocate their attentional resources to competing tasks. These include mission statements, job descriptions and monetary incentives such as bonuses (Nelton, 1994; Northcraft et al., 2011; Schmidt & DeShon, 2007). The timing and specificity of the feedback is important, and there is some research to show that if individuals are only given feedback on one task they are likely to allocate more time and effort to that task (Northcraft et al., 2011). Further, if not provided with useful high quality feedback from managers it has been shown that individuals will seek feedback elsewhere, including from the environment and coworkers, which highlights the importance of managerial feedback in encouraging accurate prioritizing of tasks (Northcraft et al., 2011).

2.5.1.4 Internet based testing.

While Internet-based testing is a relatively new phenomenon and researchers may be apprehensive about using an online format for data collection, a number of studies have revealed the benefits and advantages of online research. McGraw, Tew, and Williams (2000) collected data over a period of 2 years using an online psychology experimental laboratory. They found the quality of the data they collected to be high, the tools available online (graphics and measurement tools) to be appropriate for research purposes, and that security software helped ensure the integrity of the data. The Internet is a low cost, convenient option compared with laboratory research and provides a wider range of techniques for data collection (Brinhaum, 2004). It is an efficient means of recruiting large, heterogeneous samples (Birnhaum, 2004). Further, a number of studies (e.g., Krantz & Dalal, 2000; McGraw et al., 2000) have shown that web-based methods provide results which are similar to those obtained in the laboratory.

Certainly, the Internet has created a convenient and efficient means for the Multi-Tasks test to be administered within the workplace. As previously discussed, prior to the advent of the Internet, research in this field was confined to laboratories equipped with sophisticated technology, such as flight training centres, or professionals such as acoustic technicians (North & Gopher, 1976; Stankov& Horn, 1980). The Internet allows for presentation of the stimulus in a visually appealing and interesting way. It also allows for various manipulations of the look and timing of presentation of visual stimulus without the need for additional equipment. Competing tasks presented in this manner may prove to be more interesting than single tasks, as more information is being presented and thus must be processed. With its novel approach to competing tasks, the present study may also provide interest through engagement with the participant, which may result in stronger associations between competing tasks and intelligence (Marr & Sternberg, 1986). Numerous studies have demonstrated that response to novelty is an important aspect of intelligence and that novel stimuli can draw on abilities associated with higher intelligence, such as differential attention and ability to sort relevant from irrelevant information (Berg & Sternberg, 1984; Davidson & Sternberg, 1984; Fagan & McGrath, 1981; Lewis & Brooks-Gunn, 1981; Sternberg, 1982), thus novelty was an important consideration when developing Multi-Tasks. In addition to being novel, visually appealing and complex, Multi-Tasks was designed to be widely applicable to a range of job types and occupational settings, rather than specific to one job role. As previously discussed it therefore exploits skills known to be broadly related to job performance, including WM, Gf and attention.

König, Bühner and Mürling (2005) used a computerized measure in their multi-task study. The measure, “Simultaneous capacity/Multi-tasking” (SIMKAP), comprised five parts, the first four introducing the individual tasks and the fifth being the multi-tasking component. The tasks were: comparing stimuli and identifying the matching digits, letters or figures presented in a different field; responding to auditory questions presented by headset (logical-numerical, arithmetic and logical-verbal questions) and answering ‘work-based’ questions such as availability for lunch on a certain date. In the multi-tasking scenario these tasks were presented simultaneously (König et al., 2005). They advised that being computerised, SIMKAP had several advantages, including minimal input from the examiners, standardised administration, automated data collection and a design that could be easily replicated in future research, aiding comparisons between studies, which to date has been lacking. While

König et al. (2005)'s study used university students, SIMKAP performance was found to be related to supervisory ratings in an earlier study of Swedish navy personnel (Rosmark, 2001, as cited in König et al., 2005). However, as it has an auditory component, it requires the use of headphones, which can be difficult to arrange particularly with large groups of people in an organisational setting.

Web-based research does have its disadvantages. These include higher dropout rates and the potential for the same participant to respond twice (or more often), although cases of this are rare and reasonably easy to detect (Birnbaum, 2004). Dropouts are most likely in anonymous research or research where the individual does not benefit. A person knowledgeable about computers is also required in order to set up the website. Overall, research shows that the advantages outweigh the disadvantages of online research (Birnbaum, 2004), particularly when the researchers have extensive experience with online testing, as is the case in the current study.

2.6. Summary and rationale

Competing tasks have a long history in psychological research and particularly the intelligence literature. While many different theories have been explored as to the underlying factor of intelligence, the important feature is that all of these theories are in agreement that competing tasks are associated with performance and intelligence. There is also significant evidence to indicate that competing tasks have a stronger relationship with performance and intelligence than single tasks. Competing tasks have also been associated with individual differences in specific components of intelligence including Gf, WM and attention.

However, while competing tasks have been found to relate to intelligence and its components it is also possible that they may provide a new type of testing framework, which reveals new information about the cognitive processes underlying intelligence. That is, performance may be due to a 'competing task' factor of intelligence, or to competing tasks tapping a combination of different cognitive ability factors. Alternatively, competing tasks may measure the same underlying abilities as measures of Gf but do so in novel and challenging ways that are particularly relevant to the complex world of work.

Research to explore whether competing tasks tests: a) are relevant to the complex world of work and b) provide a new type of testing framework or measure the same underlying abilities as measures of Gf is in its infancy. This is because the administration of competing tasks has traditionally been limited by the need for specialised technology only available in laboratories or certain work contexts. This is no longer the case as the Internet has provided a convenient, efficient and widely available means for administering the competing tasks methodology to a wide range of individuals. Studies that have investigated the competing tasks methodology have shown that it is a methodology that shows promise as a useful predictor of job performance, but have been limited by the use of non-work contexts or indirect measures of job performance.

This series of studies investigated Multi-Tasks, a competing tasks measure, within a number of naturalistic work settings, compared it with other cognitive ability measures and other measures widely used in job selection settings, including measures of job performance, and investigated the underlying structure of Multi-Tasks in an effort to further develop this area of research. The components of Multi-Tasks, Word Recall and Placement Keeping were administered with a short interval between them and they ran in a continuous pattern so as to be ‘competing’. The rationale for the choice of the Word Recall and Placement Keeping tasks were firstly that they both used the visual modality and as previously discussed using the same modality results in a higher cognitive load (Brünken et al., 2002). The auditory modality was avoided as this would have required additional equipment (e.g., headphones) and would make the administration more difficult. The Placement Keeping task involved counting a series of ball movements (see Method section for further detail), thereby tapping WM and the Word Recall task was both a WM task (requiring recall of words in a certain order) and a language based task, allowing for the possibility of processing the second part of the task through different mechanisms and therefore not 'overloading' processing resources. Specific aims are given in the following section.

2.7. Study Aims

One aim of this initial two-part study was to explore a new measure of cognitive ability, the Multi-Tasks test, which employs a competing tasks methodology and the relationship between this measure and other well-known measures used in personnel selection, including measures of Gf and Gc. The Multi-Tasks test was inspired by the work of Stankov, Fogarty and colleagues (Fogarty & Stankov, 1982; Roberts et al., 1991; Stankov, 2003), drew on the CHC Theory and on the tasks used in the Woodcock Johnson III Tests of Cognitive Abilities, a psychometric measure of intelligence developed based on the CHC Theory (McGrew & Woodcock, 2001), and capitalised on modern technology, including the Internet in its design and presentation. A second and more important aim was to explore the relationship between the Multi-Tasks test and job performance. Measures of Gf and Gc are commonly used in psychometric testing for job selection because they measure constructs known to be important for job performance, including communication, retention of information, and reasoning ability. Gs has been included in the design as it is an important component of the CHC theory and mental speed has been prominent (though somewhat controversial) in the intelligence literature. Research on mental speed as a predictor of job performance is meager, however, and Gs is not expected to play a significant role in predicting job performance.

Specific aims of the first two-part study were to: (1) determine the general properties, including the reliabilities and means of the new test (Part A (the pilot study); (2) assess the predictive validity of competing tasks; (3) determine whether the test measures the broad factors described by the CHC theory; and (4) investigate the relationship between competing task performance and job performance. An additional aim was to assess differences among various demographic groups in performance on the test.

Three theoretical models that follow were proposed to address the study aims. These are outlined in further detail in the Results. The first was the proposed model for the competing tasks measure (Multi-Tasks). It was hypothesised that the underlying dimensions of the competing task measure would include factors relating to the two tasks that comprised the measure and a factor relating to the speed at which the tasks were undertaken. This is depicted in Figure 1.2.

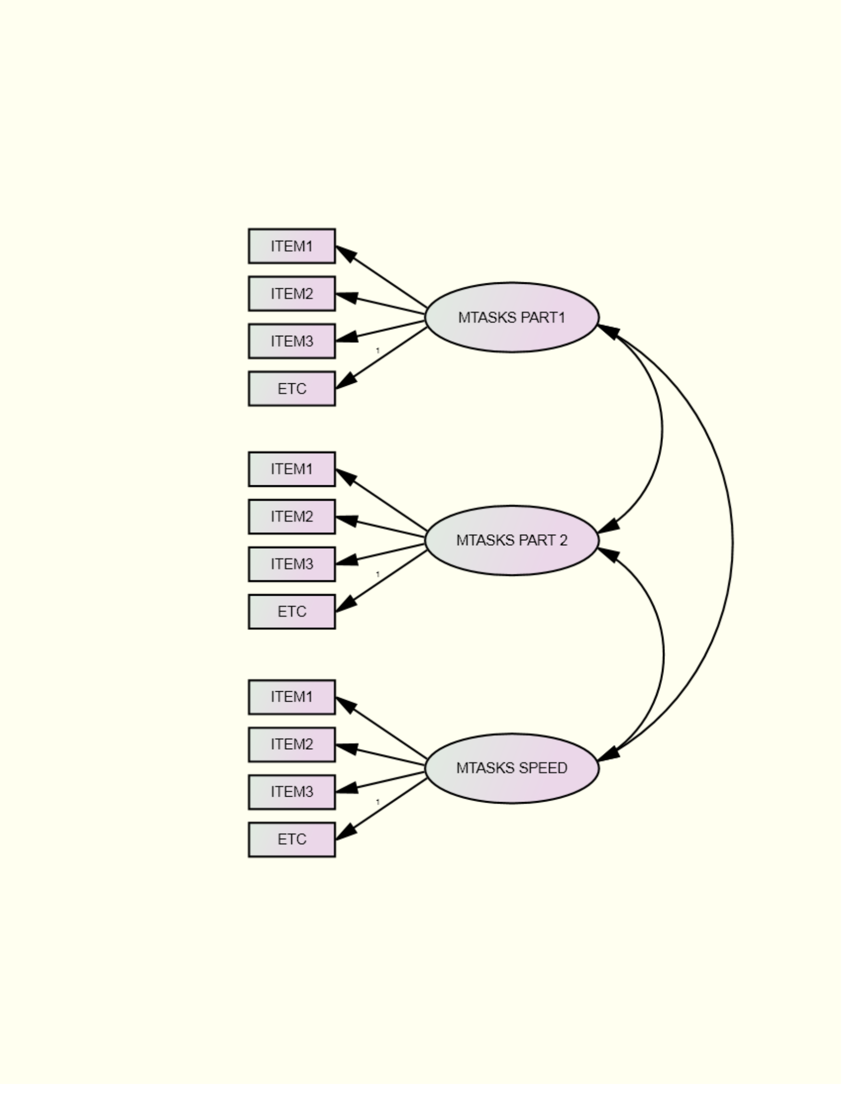


Figure 1.2. Proposed Multi-Tasks model.

The second theoretical model, depicted in Figure 1.3 addresses the relationship between the competing tasks measure and other well-known measures used in job selection and the CHC theory of intelligence. It was hypothesised that the components of the Multi-Tasks measure would be strongly associated with Gf (König, et al., 2005; Stankov, 1988), as would the task used as a criterion measure of Gf (Digit Sequence), whereas the measure included as a criterion measure of Gc (Word Reasoning) would load onto the Gc factor.

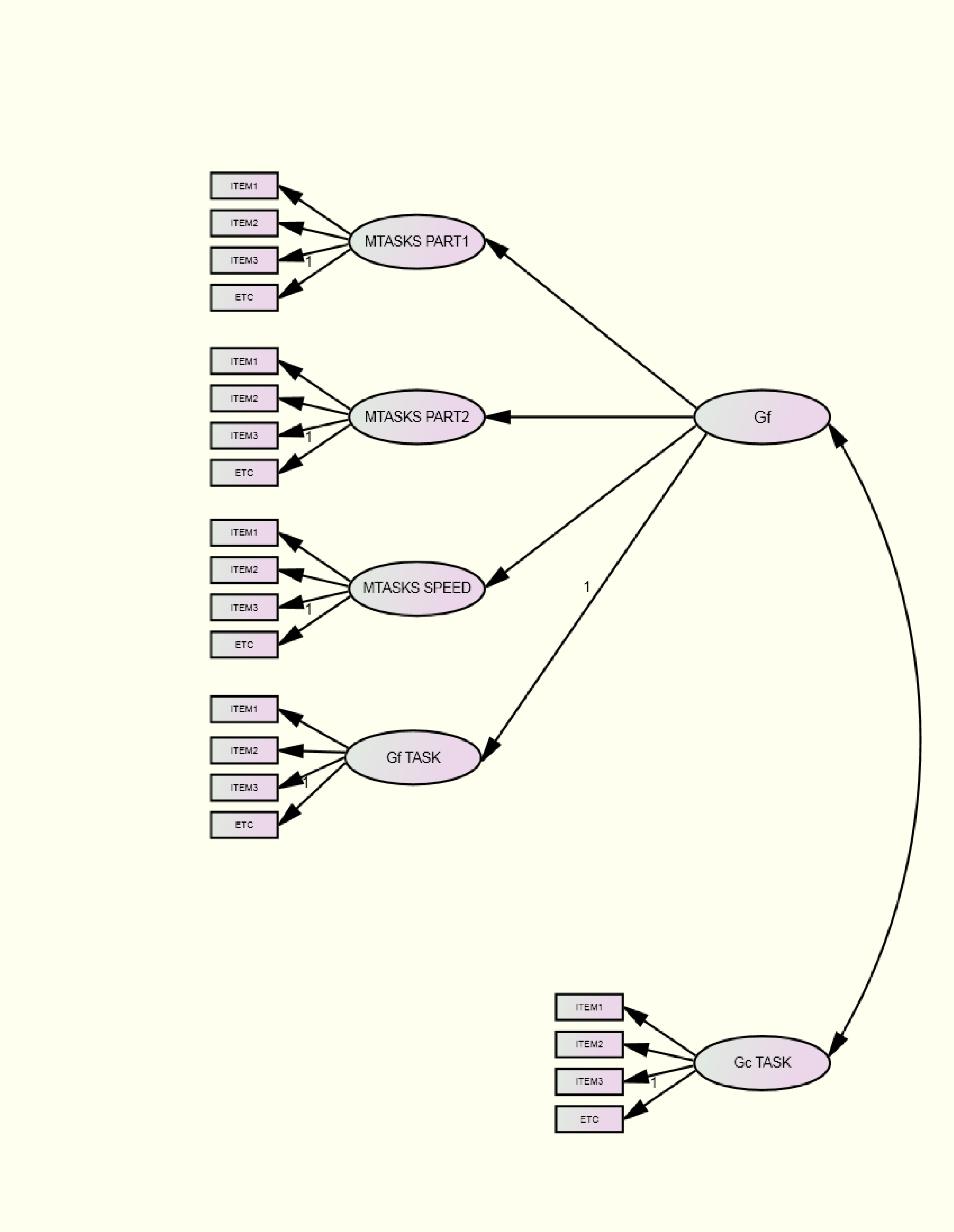


Figure 1.3. Proposed model for Multi-Tasks, Gf, and Gc measures.

The final theoretical model depicted in Figure 1.4 outlined the proposed relationship between all the tasks and the job performance measures. It was hypothesised that all tasks would predict all components of job performance, however, the relationship between the Multi-Tasks test and job performance would be stronger than the other intelligence measures.

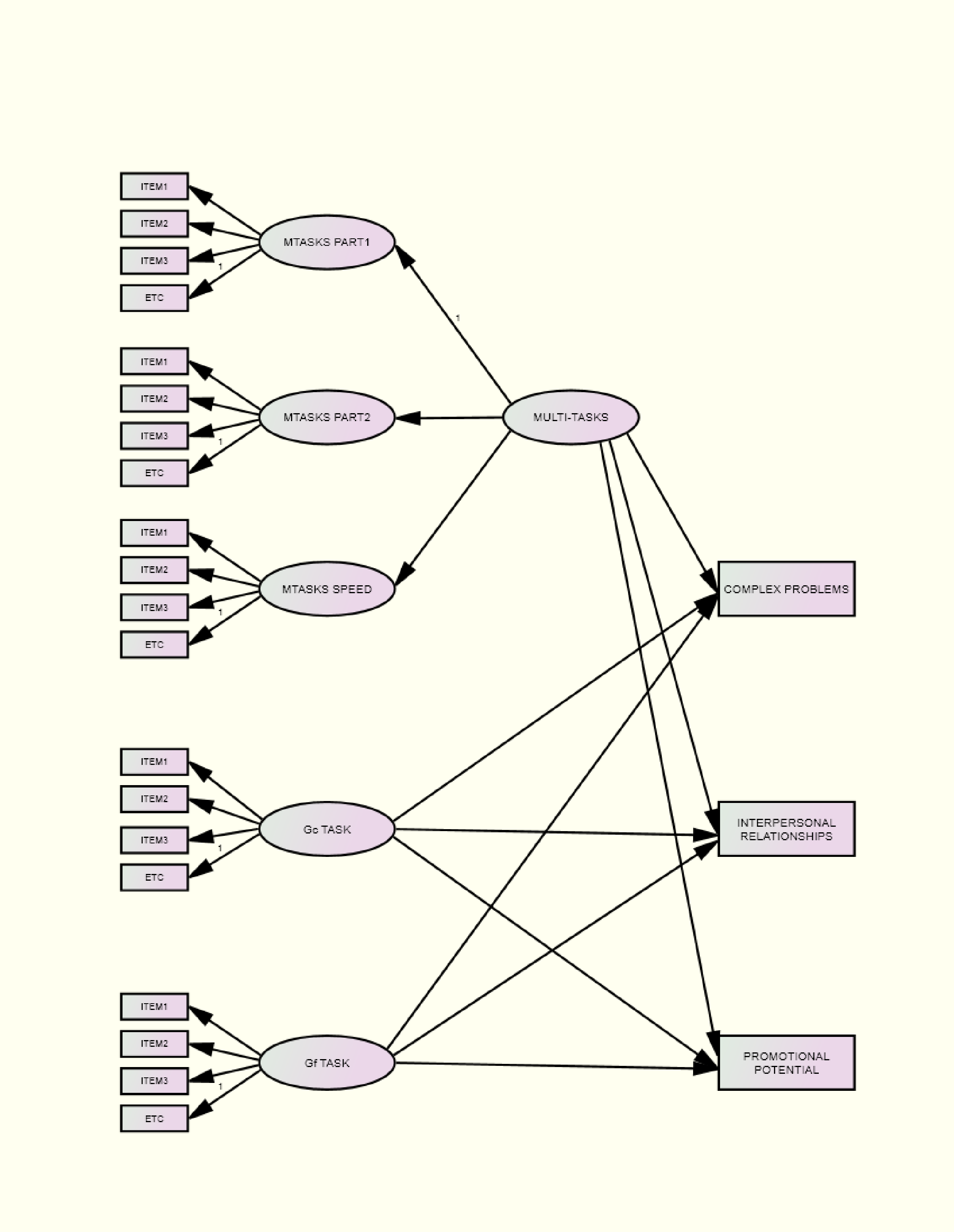


Figure 1.4: Proposed model testing the predictive validity of the intelligence measures against the job performance measures.

PART A (PILOT STUDY)

2.8 Method

2.8.1 Participants.

A total of 148 job applicants for customer service positions within the financial services industry in Australia participated in the study. The group consisted of 94 (64%) males and 54 (36%) females. Age and education level of the participants are shown in Tables 1.2 and 1.3 respectively.

Table 1.2

Age of participants (N = 148)

Variable	Frequency	Percentage
Age		
18 to 24	32	21.6
25 to 29	71	48.0
30 to 34	24	16.2
35 to 39	4	2.7
40 to 44	17	11.5

Table 1.3

Education level of participants (N = 148)

Variable	Frequency	Percentage
Education level		
Secondary	6	4.1
Higher school certificate	32	21.6
Tafe qualified	32	21.6
Took college courses	2	1.4
Undergraduate university degree	49	33.1
Postgraduate university degree	24	16.2
Missing	3	2.0

2.8.2 Materials.

The study employed a new experimental cognitive ability measure called the Multi-Tasks test. The Multi-Tasks test was designed to be engaging and visually appealing yet not overly complex so as to require a large degree of explanation. For this reason it used one novel task that was fluid and interesting (Placement Keeping) and one task that was less engaging but likely to be somewhat familiar as it has been widely used in previous studies and occupational settings (Word Recall) (Stankov, 2003; Wang, Proctor & Pick, 2007). Consideration was also given to making it user-friendly, widely applicable to a range of jobs and not susceptible to technical issues (such as difficulties coordinating visual and sound stimuli) or occupational health and safety issues (such as the need for sharing headphones) and able to be deployed easily for large groups of people from any organisation (including cross-culturally) simultaneously. It was therefore Internet based and in the visual modality only. There was a slight delay between the presentations of the two tasks so that it was possible to attend to both tasks. The Multi-Tasks test consisted of two sub-tasks presented simultaneously, with each sub-task being composed of 16 items. The first sub-task is referred to as “Placement Keeping”. Each item of this sub-task consisted of balls moving amongst three boxes. Following a sequence of ball movements, the participant was asked to identify how many balls remain in each box. A drop-down menu prompted the participants to select a response, from 12 alternatives. The test included two practice questions, a 5-second familiarisation time for each new question, and a 1.5-second delay between ball movements. Each item was scored for the total number of boxes correct, with a minimum of 0 and a maximum of 3 for each item, providing a maximum score of 48 for this sub-task. The number of ball movements increased as the questions progressed. The practice questions and test questions 1 to 4 had four ball movements, questions 5 to 8 had six ball movements, questions 9 to 12 had eight and questions 13 to 16 had ten. It was hypothesised that this would measure a construct similar to Gf from the CHC Theory.

The second sub-task, “Word Recall”, involved participants being presented with a word following each ball movement. The word was displayed for 1.5 seconds following each ball movement until the next ball movement commenced. After the presentation of the item, the participant was asked to select from a drop-down menu of ten alternatives, the ordinal position in which a particular target word appeared. Each item was scored 0 for incorrect or 1 for correct, making a maximum score of 16 for this sub-task. The number of words presented matched the number of ball movements, so that the practice questions and test questions 1 to 4 had four words presented, questions 5 to 8 had six words, questions 9 to 12 had eight and questions 13 to 16 had ten. It was hypothesised that this would measure a construct similar to Gf or Gc from the CHC Theory.

Thus, overall scores on the Multi-Tasks test could range from 0 to 64. In addition to this composite score, a measure reflecting the speed of test-taking was also collected. This

"Broad Speediness" score was based on the time taken to complete an item from the time of presentation to response. The time taken was reported in milliseconds. There was a time limit of 60 seconds per question. This was hypothesised to be a measure of Gs.

Figure 1.5 illustrates the Multi-Tasks subscales.

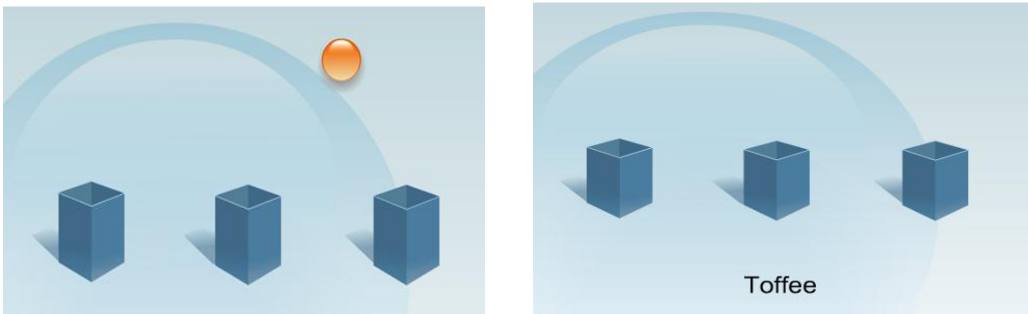


Figure 1.5. Multi-Tasks test Placement Keeping sub-task (left), then Word Recall sub-task (right).

2.8.3 Procedure.

Testing was undertaken on computer at the site of the workplaces involved in the study. Each participant undertook the assessment individually. An invigilator from the workplace greeted each participant and assisted him or her to login to the system. The system was web-based and supported by Microsoft.net and Adobe Flash on a Microsoft Sequel Server. The system provided the participants with full instructions, including practice questions. Each participant was administered the same version of Multi-Tasks to ensure testing was standardised across all participants. Testing time was approximately 12 minutes. When the tests were completed the system informed the participant that they had completed the assessment. Results were saved to a secure server. Outcomes were analysed using the IBM Statistical Package for the Social Sciences.

2.9 Results

In the first section, preliminary procedures are described. In particular, the outlier checks, reliabilities of the measures, and descriptive statistics are reported. Following this, t-test and analysis of variance results capturing demographic group differences in Multi-Tasks test scores are described.

2.9.1 Preliminary Procedures.

2.9.1.1 Outlier checks.

Total scores were computed for the Word Recall and Placement Keeping Multi-Tasks subscales as well as completion time. All measures were standardised; cases where standardised values exceeded the absolute value of 3.29 (per Tabachnick & Fidell, 2007) were considered as outliers. No cases were found to have standardised values higher than the absolute value of 3.29; therefore, no cases were deleted from subsequent analyses.

2.9.1.2 Descriptive statistics and reliabilities of the measures.

The descriptive statistics for the variables are displayed in Table 1.4. The Multi-Tasks Word Recall subscale was reliable (α = .785) (Hair, Black, Babin, & Anderson, 2010). The mean Word Recall score was 8.58 (SD = 3.46) out of a total possible score of 16.

The Multi-Tasks Placement Keeping subscale had high internal consistency (α = .841). The mean Placement Keeping score was 29.14 (SD = 10.04) out of a total possible score of 48. The alpha for Broad Speediness was .939 and the mean score in seconds was 5.34 (SD = 2.19).

Table 1.4

Descriptive statistics and alpha coefficients for the variables

Subscale	Case N	Item N	M	SD	Alpha
Multi-Tasks					
Word recall	148	16	8.58	3.46	.785
Placement keeping	148	16	29.14	10.04	.841
Broad Speediness	148	16	5.34	2.19	.939

2.9.1.3 Multi-Tasks scores across demographic groups.

Independent-samples t-test and one-way analysis of variance (ANOVA) procedures were conducted to determine whether scores vary across gender, levels of education, age, and first language. A two-tailed significance level was specified, and a significance level of .05 was also specified.

2.9.1.3.1 Gender.

The means and standard deviations for the Multi-Task tests are presented in the following table. Results indicate that the Multi-Tasks scores of males did not significantly differ from that of females.

Table 1.5

Means and standard deviations for Multi-Tasks scores of males and females

Gender	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Males	94	8.89	3.56	30.07	9.08
Females	54	8.04	3.24	27.50	11.42

2.9.1.3.2 Level of education.

The means and standard deviations for the Multi-Task tests are displayed in the following table. The findings reveal that Multi-Tasks Word Recall scores did not vary significantly across levels of education. However, Multi-Tasks Placement Keeping scores did differ significantly across levels of education ($F(6,141) = 4.63$, $p < .001$). Individuals with higher levels of education tended to do better here.

Table 1.6

Means and standard deviations for Multi-Tasks scores across levels of education

Level of Education	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Secondary	6	7.83	3.37	25.33	10.39
Higher school certificate	32	7.88	3.53	25.59	10.32
TAFE qualified	32	8.28	3.75	27.16	10.90
College courses	2	7.50	6.36	27.50	.71
Undergraduate	49	9.14	3.33	31.04	9.37
Postgraduate	24	9.42	3.02	35.54	4.69

2.9.1.3.3 Age.

The means and standard deviations for the Multi-Tasks tests are shown in the following table. No significant differences in Multi-Tasks were found on the basis of age.

Table 1.7

Means and standard deviations for Multi-Tasks scores across age groups

Age Group	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
18 to 24	32	8.84	3.15	30.00	10.32
25 to 29	71	8.79	3.64	29.89	9.70
30 to 34	24	7.88	3.43	27.00	10.57
35 to 39	4	10.50	2.89	29.75	2.06
40 to 44	17	7.76	3.44	27.24	11.46

PART B

2.10. Method

2.10.1 Participants.

A total of 512 job applicants for a range of positions within business and telecommunication services in Australia participated in the study. The group consisted of 269 (54.1%) males and 229 (45.9%) females. Fourteen participants did not specify their gender. Age and education level of the participants are shown in Tables 1.8 and 1.9 respectively.

Table 1.8

Age of participants (N = 512)

Variable	Frequency	Percentage
Age		
18 to 24	108	21.1
25 to 29	87	17.0
30 to 34	68	13.3
35 to 39	63	12.3
40 to 44	56	10.9
45 to 49	59	11.5
50 to 54	33	6.4
55 and above	23	4.5
Missing	15	2.9

Table 1.9

Education level of participants (N = 512)

Variable	Frequency	Percentage
Education level		
Secondary certificate	23	4.5
Higher school certificate	53	10.4
Tafe qualified	129	25.2
Took college courses	144	28.1
Undergraduate university degree	104	20.3
Postgraduate university degree	43	8.4
Other	13	2.5
Missing	3	.6

2.10.2 Materials.

The study employed three measures of cognitive ability and three measures of job performance. The ability measures included two tests that were used as criterion measures of fluid and crystallized intelligence. These tests are called the Digit Sequence test and Word Reasoning test. The third test was the new experimental measure called the Multi-Tasks test. The measures of job performance consisted of supervisory ratings. Each of these tests will now be described in more detail.

2.10.2.1 Digit Sequence test.

In the Digit Sequence test (source: E-ntelligent Testing Products Pty Ltd), participants are given a block of five, four digit-long items, which, with the exception of one, all follow predetermined numerical rules. The participants' task was to determine which among the five is the odd one out and to click on it using a computer mouse. For instance, all but one of the blocks of digits may have had a 4 as the second number. In this instance, the odd one out would be the one which did not have a 4 as the second number and the participant should click that number block. There were 12 items in total in the test. A variant of this test, known as Letter Sets, comes from the work of Ekstrom, French, and Harman (1979), who have shown it to be a measure of the Induction primary factor and fluid intelligence (Gf) at the second-order. Responses were scored 0 for incorrect and 1 for correct for each item, leading to a total "Number Correct" score out of 12. This was the accuracy score. A second score, measuring the speed of the response in milliseconds, was also recorded. There was a time limit of 60 seconds for each question. While some authors have argued against imposing time limits on intelligence measures (e.g., Carroll, 1993; Wilhelm & Schulze, 2002), time limits are necessary in job selection research as it is not practical to allow applicants to spend infinite time on tasks and it may provide the opportunity for them to obtain assistance or undertake research, which would invalidate the measures. Measures were taken of the time to respond for all 12 items and then a final score was computed which was the average of these 12 scores. The dependent variable in this case was time taken overall. Figure 1.6 provides an illustration of a Digit Sequence item.

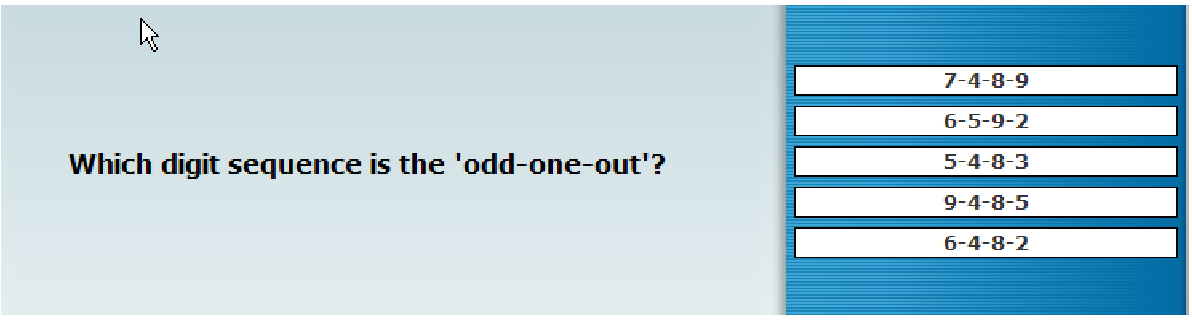


Figure 1.6. Sample of a digit sequence test item.

2.10.2.2 Word Reasoning test.

In the Word Reasoning test (source: E-ntelligent Testing Products Pty Ltd), designed to assess the Cognition of Semantic Relations primary mental ability, and crystallized intelligence (Gc) at the second-order, participants were asked to select, from among five alternatives, the option which completes a given verbal analogy. Responses were scored 0 for incorrect and 1 for correct for each item, the total being the accuracy score out of a possible 22. As per the Digit Sequence Test as a speed of response score was also obtained, (again measured in milliseconds) for each question and then an overall score was computed from the average of these 22 scores. There was a time limit of 45 seconds for each question. Figure 1.7 provides an illustration of a Word Reasoning item.

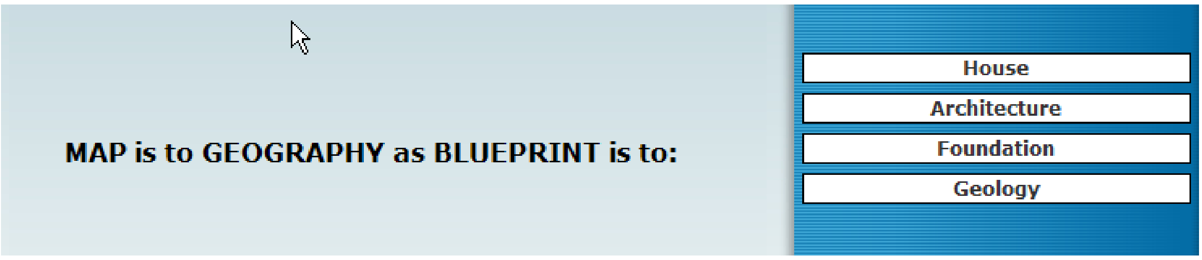


Figure 1.7. Sample of word reasoning test item.

2.10.2.3 Multi-Tasks test.

For all full description of the Multi-Tasks test see Study 1 (Part A). A video demonstration is also provided at: http://rightpeople.com.au/researchvideos/ (username: RightPeople, password: Rp1111117). Please visit the link now. Copy and paste the URL into your Internet browser rather than clicking on it. The number of balls is displayed in front of each box momentarily before the presentation begins. Use the pause button to halt the presentation if required.

2.10.2.4 Job performance measures.

Supervisors of applicants who had been hired and employed for six months rated applicants on three items, including (1) Ability to solve complex tasks; (2) Interpersonal skills and ability; and (3) Promotional potential. Each aspect of job performance was rated on a 5 point scale either as 1 (Unsatisfactory, meaning the person is consistently below expectations), 2 (Improvement needed, meaning the person does not consistently meet expectations), 3 (Meets expectations, meaning the person consistently meets expectations), 4 (Exceeds expectations, meaning the person consistently exceeds expectations), or 5 (Exceptional, meaning the person far exceeds expectations). Ratings were collected by Human Resource Officers and provided to the examiners on an excel spreadsheet. Ratings were able to be obtained for 85 participants.

2.10.3 Procedure.

Testing was undertaken on computer at the site of the workplaces involved in the study. Each participant undertook the assessment individually. An invigilator from the workplace greeted each participant and assisted him or her to login to the system. The system was web-based and supported by Microsoft.net and Adobe Flash on a Microsoft Sequel Server. The system provided the participants with full instructions, including practice questions (2 questions per sub-test) and feedback if the practice questions were completed incorrectly. Each participant was administered the test battery with the same order of sub-tests. The test order was Word Reasoning, Digit Sequence and then Multi-Tasks. Thus, testing was standardised across all participants. Testing time was approximately 45 minutes. When the tests were completed the system informed the participant that they had completed the assessment. Results were saved to a secure server. Outcomes were analysed using the IBM Statistical Package for the Social Sciences including AMOS for structural equation modelling.

2.11. Results

In the first section, preliminary procedures will be described. In particular, the outlier checks, reliabilities of the measures, and descriptive statistics will be reported. Following this, the results of the theoretical measurement model and structural equation model tests will be presented. In the last section, t-test and analysis of variance results capturing demographic group differences in Multi-Tasks test scores will be described.

2.11.1 Preliminary Procedures.

2.11.1.1 Outlier checks.

Total scores were computed for the Word Recall and Placement Keeping Multi-Tasks subscales while mean composites were computed for the Broad Speediness Multi-Task measure. The composites were standardised; cases where standardised values exceeded the absolute value of 3.29 (per Tabachnick & Fidell, 2007) were considered as outliers. Fourteen cases had standardised values higher than the absolute value of 3.29 and were deleted from subsequent analyses.

2.11.1.2 Descriptive statistics and reliabilities of the measures.

The descriptive statistics for the variables are displayed in Table 1.10. The Multi-Tasks Word Recall subscale was reliable ($\alpha = .86$) (Hair, Black, Babin & Anderson, 2010). The mean Word Recall score was 8.24 (SD = 3.42).

The Multi-Tasks Placement Keeping subscale had high internal consistency ($\alpha = .93$). The mean Placement Keeping score was 21.56 (SD = 10.96). The alpha for Digit Sequence was .59 and the mean score was 7.28 (SD = 1.89). The alpha for Word Reasoning was .75 and the mean score was 12.96 (SD = 3.60).

The mean scores for the three outcome variables – Complex Problems, Interpersonal Problems and Promotion Potential ranged from 3.47 to 3.59.

Table 1.10

Descriptive statistics and alpha coefficients for the variables

Subscale	Case N	Item N	M	SD	Alpha
Multi-Tasks					
Word recall	498	12	8.24	3.42	.86
Placement keeping	498	12	21.56	10.96	.93
Broad speediness (in milliseconds)	498	12	31354.79	13276.22	.92
Digit Sequence					
Number Correct	498	10	7.28	1.89	.59
Speed (in milliseconds)	498	10	33857.92	11169.33	.70
Word Reasoning					
Number Correct	498	20	12.96	3.60	.75
Speed (in milliseconds)	498	20	13169.27	4339.75	.87
Complex Problems	85	--	3.59	.79	--
Interpersonal Problems	85	--	3.47	.61	--
Promotional Potential	85	--	3.58	.70	--

2.11.1.3 Correlations among variables.

Since most of the variables were not normally distributed (cf. the histograms in Appendix 1A), Kendall's tau correlations between the variables are reported in Table 1.11. The findings reveal that as expected, the Multi-Tasks test components were positively correlated with the job performance measures. Multi-Tasks Word Recall subscale was positively correlated with the Digit Sequence scale ($r = .08$, $p < .05$), Complex Problems (r =

.41, $p < .01$), and Promotional Potential ($r = .35$, $p < .01$). The Multi-Tasks Placement Keeping subscale was positively correlated with the Digit Sequence scale ($r = .30$, $p < .01$), Word Reasoning ($r = .28$, $p < .001$), time it took to answer the Word Reasoning items (Word Reasoning speed) ($r = .09$, $p < .01$), Complex Problems ($r = .52$, $p < .001$), and Promotional Potential ($r = .44$, $p < .001$). The time it took to complete the Multi-Tasks Word Recall and Placement Keeping tasks (Multi-Tasks speed) was positively correlated with the Digit Sequence scale ($r = .12$, $p < .001$), Digit Sequence speed ($r = .20$, $p < .001$), the Word Reasoning scale ($r = .15$, $p < .001$) and Word Reasoning speed ($r = .19$, $p < .001$).

Multi-Tasks speed was positively correlated with Complex Problems ($r = .21$, $p < .05$) and Promotional Potential ($r = .20$, $p < .05$), but not with Interpersonal Problems ($r = .06$, $p > .05$). Word Reasoning and Digit Sequence speed measures were not correlated with any of the job performance measures.

It is also noteworthy that Complex Problems and Promotional Potential were highly correlated ($r = .88$, $p < .001$), which suggests that respondents were not distinguished by supervisors between these two types of performance in any meaningful way, and that ability to solve complex tasks is associated with better job performance, as discussed in Section 2.5.

Table 1.11

Kendall Tau correlations between variables (N = 498)

Variable	1	2	3	4	5	6	7	8	9
1 Multi-Tasks word recall									
2 Multi-Tasks placement keeping	.32***								
3 Multi-Tasks speed	.23***	.22***							
4 Digit Sequence number correct	.08*	.30***	.12***						
5 Digit Sequence speed	-.04	.02	.20***	.01					
6 Word Reasoning number correct	.05	.28***	.15***	.33***	.06				
7 Word Reasoning speed	.02	.09*	.19***	.14***	.31***	.03			
8 Complex Problems[1]	.41***	.52***	.21**	.21**	-.02	.20*	.04		
9 Interpersonal Problems[1]	-.06	.02	.06	.13	.05	.40***	-.05	.11	
10 Promotional Potential[1]	.35***	.44***	.20*	.18	-.00	.22*	.05	.88***	.16

[1] N = 85.

*p < .05 **p < .01 ***p < .001

2.11.1.4 Multi-Tasks scores across demographic groups.

Independent t-test and one-way analysis of variance (ANOVA) procedures were conducted to determine whether scores would vary across gender, levels of education, and age. A two-tailed significance level was specified for the independent t-test procedure. A significance level of .05 was specified for the overall model in the ANOVA procedure; when the overall model was statistically significant, post-hoc Tukey procedures were conducted.

2.11.1.4.1 Gender.

The means and standard deviations for the Multi-Task tests are presented in Table 1.12. Results indicate that the Multi-Tasks Word Recall scores of males did not significantly differ from that of females. Males (M = 23.92, SD = 9.82; t (448) = .40, NS; r^2 = .000) tended to have higher Multi-Task Placement Keeping scores than females (M = 18.85, SD = 11.63; t (448) = 5.21, p <.01; r^2 = .057).

Table 1.12

Means and standard deviations for Multi-Tasks scores of males and females

Gender	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Males	269	8.32	3.22	23.92	9.82
Females	229	8.20	3.59	18.85	11.63

2.11.1.4.2 Level of education.

The means and standard deviations for the Multi-Task tests are displayed in Table 1.13. The findings reveal that Multi-Tasks Word Recall scores did not vary significantly across levels of education (F (5,490) = 1.98, p > .05; η^2 = .020). Similarly, Multi-Tasks Placement Keeping scores did not differ significantly across levels of education (F (5,490) = .89, p > .05; η^2 = .001).

Table 1.13

Means and standard deviations for Multi-Tasks scores across levels of education

Level of Education	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Secondary certificate	23	8.91	3.16	23.13	10.42
Higher school certificate	53	9.36	3.25	20.00	12.11
TAFE qualified	129	8.45	3.09	21.44	10.62
Took college courses	144	7.93	3.54	20.82	11.23
Undergraduate degree	104	7.84	3.60	23.11	10.48
Postgraduate degree	43	8.33	3.29	22.30	11.42

2.11.1.4.3 Age.

The means and standard deviations for the Multi-Tasks tests are shown in Table 1.14. The findings indicate that Multi-Tasks Word Recall scores varied across age groups (F (7,489) = 2.60, p <.05; η^2 = .036). Post-hoc Tukey tests reveal that participants between 25 to 29 years had significantly lower Word Recall scores (M = 7.29, SD = 3.54) than participants between 35 and 39 years (M = 9.00, SD = 3.01; p <.05) and participants between 40 and 44 years of age (M = 9.41, SD = 2.97; p <.01).

Multi-Tasks Placement Keeping scores also varied significantly across age groups (F (7,489) = 3.58, p <.01; η^2 = .049). Post-hoc Tukey test procedures reveal that participants between 40 and 44 years had significantly higher Placement Keeping scores (M = 26.36, SD = 9.74) than participants between 18 and 24 years (M = 20.64, SD = 10.96; p < .05), participants between 25 and 29 years (M = 20.01, SD = 10.97; p <.05), and participants between 30 and 34 years of age (M = 17.88, SD = 11.61; p <.001).

Table 1.14

Means and standard deviations for Multi-Tasks scores across age groups

Age Group	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
18 to 24	108	8.25	3.41	20.64	10.96
25 to 29	87	7.29	3.54	20.01	10.97
30 to 34	68	8.18	3.56	17.88	11.61
35 to 39	63	9.00	3.01	21.95	11.54
40 to 44	56	9.41	2.97	26.36	9.74
45 to 49	59	7.95	3.42	22.77	10.25
50 to 54	33	8.18	2.90	23.88	8.01
55 and above	23	8.83	3.87	24.04	11.97

2.11.2 Results of Measurement and Structural Model Tests.

This section outlines the results of the measurement model and structural equation model tests after a discussion regarding how the data was treated in the analysis.

2.11.2.1 Parcels for measurement and structural model tests.

Parcels (i.e., small item groups) were created for two reasons. Firstly, the tests were scored using a nominal scale (i.e., incorrect vs. correct), however, one requirement of structural equation modeling is that the data be measured on either an interval or ratio scale. Secondly, as Little, Cunningham, Shahar, and Widaman (2002) indicated, models with single-item indicators are less parsimonious and often increase sampling error.

2.11.2.2 Multi-Task parcels.

The Multi-Tasks items were arranged in order of difficulty. Parcels were thus created based on number order. Four parcels consisting of three items were created; the first parcel had the least difficult items while the fourth parcel had the most difficult items (refer to Appendix 1B for the parcel compositions).

2.11.2.3 Digit sequence and word reasoning parcels.

Since items for these tests were not arranged in order of difficulty, the item-to-construct balance method, recommended by Little et al. (2002), was used to create parcels. Specifically, the corrected item-total correlations (from the reliability analyses) were used to generate the parcels. These correlations were sorted from highest to lowest. Items with the highest item-total correlation anchored each of the parcels. The items with the next highest item-total correlations were added to the anchors in the reverse order. The item with the highest correlation among the anchor items was matched with the lowest loading item from the second selection. This basic procedure where items with lower correlations were matched with items with higher correlations was repeated until all items were categorized into parcels. Refer to Appendix 1B for the parcel compositions.

2.11.2.4 Normality of parcels.

As multivariate normality is an important assumption of structural equation modeling, normality of the parcels was assessed. According to Kline (2005), multivariate normality can be assumed when the univariate distributions are normal and the distribution of any pair of variables is bivariate normal. Kline (2005) further indicated that because it is impractical to examine all joint distributions, examining univariate distributions will usually allow one to detect instances of non-normality. Thus, the skewness and kurtosis of the parcels were assessed. Most of the variables were highly skewed; their skew indices were above the acceptable criterion of three (Kline, 2005). Accordingly, these variables were transformed to correct for skewness (Howell, 1992). The variables that were negatively skewed were normalised using a power transformation; the variables that were positively skewed were normalized using a square root transformation. These transformed variables were used in subsequent analyses.

2.11.3 Model Evaluation.

2.11.3.1 Procedure for model evaluation.

Model fit was assessed by interpreting several fit indices such as the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), and the likelihood ratio χ^2 test. A model is deemed as fitting the data well when the CFI value is above .95 (Hu & Bentler, 1999). Browne and Cudeck (1993) suggested that a model with an RMSEA value less than .05 has good fit, one with a value than .08 has reasonable fit, and a model with an RMSEA less than .10 has poor fit. Kline (2005) proposed that a model with an SRMR value of less than .10 has good fit. A small χ^2 value relative to the degrees of freedom indicates good model fit (Hu & Bentler, 1999).

Nested models were compared via the change in chi-square. In addition, two information indices were reported: the Akaike Information Criterion (AIC) and the Expected Cross-Validation Index (ECVI). Information indices are relative; models indices are compared to each other and the lower the value, the better the model fit (Byrne, 2001).

In addition to evaluating the model as a whole, the fit of the individual parameters was also assessed (Byrne, 2001). Parameters were evaluated at the .05 level. The direction of the standardized path coefficients was checked to see if it was consistent with expectations.

The model was tested in three stages: the measurement model for the Multi-Tasks test, the full measurement model and then the measurement plus the structural model.

2.11.3.2 Confirmatory factor analysis of the Multi-Tasks test.

A single-factor, two-factor (test and speed constructs), and three-factor model using the Multi-Tasks parcels were tested. The fit indices for all three theoretical models are summarised in Table 1.15. Figure 1.8 outlines the three-factor model.

The findings reveal that none of the three models fitted the data perfectly but the three-factor model came closest with some of the fit statistics meeting the criteria outlined earlier (e.g., CFI and SRMR). Further, the change in chi-square between the single-factor and three-factor model was statistically significant ($\Delta\chi^2(2) = 1748.48$, $p < .001$); the change in chi-square between the two-factor and three-factor model was also statistically significant ($\Delta\chi^2(3) = 653.46$, $p < .001$). Lastly, values of the information indices for the three-factor model were lower than the values for the single- and two-factor models. Maximum likelihood estimates for the variables, and correlations among the factors are shown in Appendix 1C. Although fit statistics for the three-factor model were marginal, this three-factor model was carried through to the next stage of the analysis before further modifications were made.

Table 1.15

Fit indices for the Multi-Tasks measurement models

Index	Single Factor	Two Factor	Three Factor
Chi-square	2007.58	912.56	259.10
Degrees of freedom	54.00	53.00	51.00
Sig.	.00	.00	.00
Chi-square/df	37.18	17.22	5.08
Comparative fit index (CFI)	.50	.78	.95
Root mean squared error (RMSEA)	.27	.18	.09
Lower bound 90 % CI	.26	.17	.08
Upper bound of 90 % CI	.28	.19	.10
Standardized root mean residual (SRMR)	.19	.13	.05
Akaike information criterion (AIC)	2055.58	962.56	313.10
Expected cross-validation index (ECVI)	4.02	1.88	.61

Note. At $p < .001$, critical χ^2_{crit} (2) = 13.82 and χ^2_{crit} (3) = 16.27.

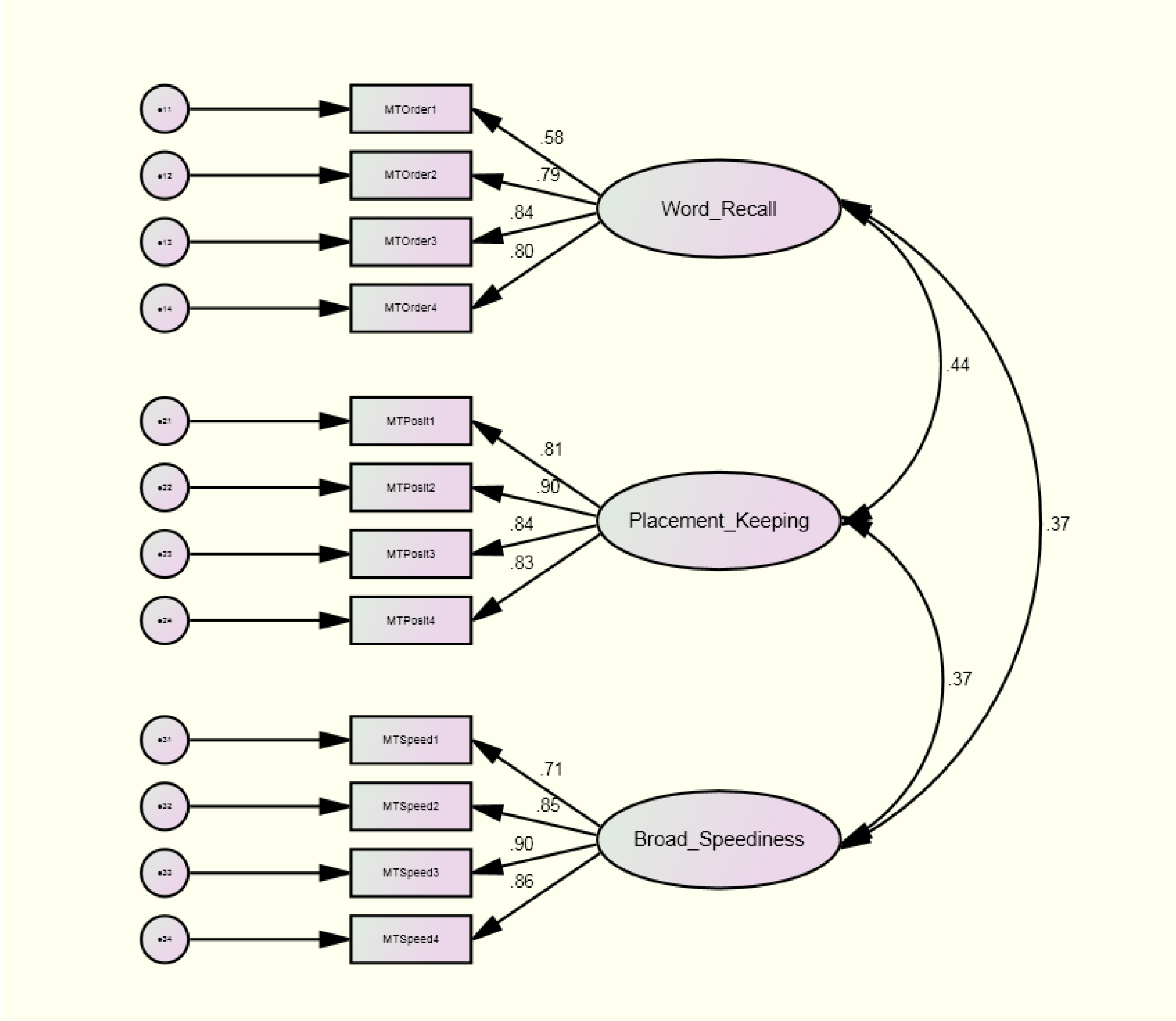


Figure 1.8. Three-factor measurement model for Multi-Tasks (standardised coefficients).

2.11.3.3 Confirmatory factor analysis of the full measurement model.

The proposed full Multi-Task measurement model is depicted in Figure 1.9. The fit indices for this model are summarised in Table 1.16. The path coefficients from the second-order construct to the first-order constructs are presented in Appendix 1D. This two-factor model fit the data well as all index values reached their respective benchmarks. As shown in Figure 1.9, Digit Sequence, the two Multi-Tasks subtests and the Multi-Tasks speed measure loaded onto the Fluid Intelligence factor while Word Reasoning represented the Crystallized Intelligence factor.

Since the standardised loading of the Multi-Tasks Broad Speediness factor was low at .42 and because the Digit Sequence and Word Reasoning speed factors were not included in the model (as measurement models with these factors had inadequate fit), the Broad Speediness construct was dropped from the model. The revised theoretical model of intelligence is depicted in Figure 1.10 and the fit indices are summarized in Table 1.16. This revised model fitted the data well. It also had significantly better fit than the proposed model: the change in chi-square was statistically significant ($\Delta\chi^2(69) = 201.47$, $p < .001$) and the values of the information indices of the revised model were lower than those of the proposed model.

Table 1.16

Fit indices for the intelligence model

Index	Proposed	Revised
Chi-square	467.49	266.02
Degrees of freedom	184.00	115.00
Sig.	.00	.00
Chi-square/df	2.54	2.31
Comparative fit index (CFI)	.94	.95
Root mean squared error (RMSEA)	.06	.05
Lower bound 90 % CI	.05	.04
Upper bound 90% CI	.06	.06
Standardized root mean residual (SRMR)	.06	.06
Akaike information criterion (AIC)	561.49	342.02
Expected cross-validation index (ECVI)	1.10	.67

Note. At $p < .001$, critical $\chi^2_{crit}(69) = 111.06$.

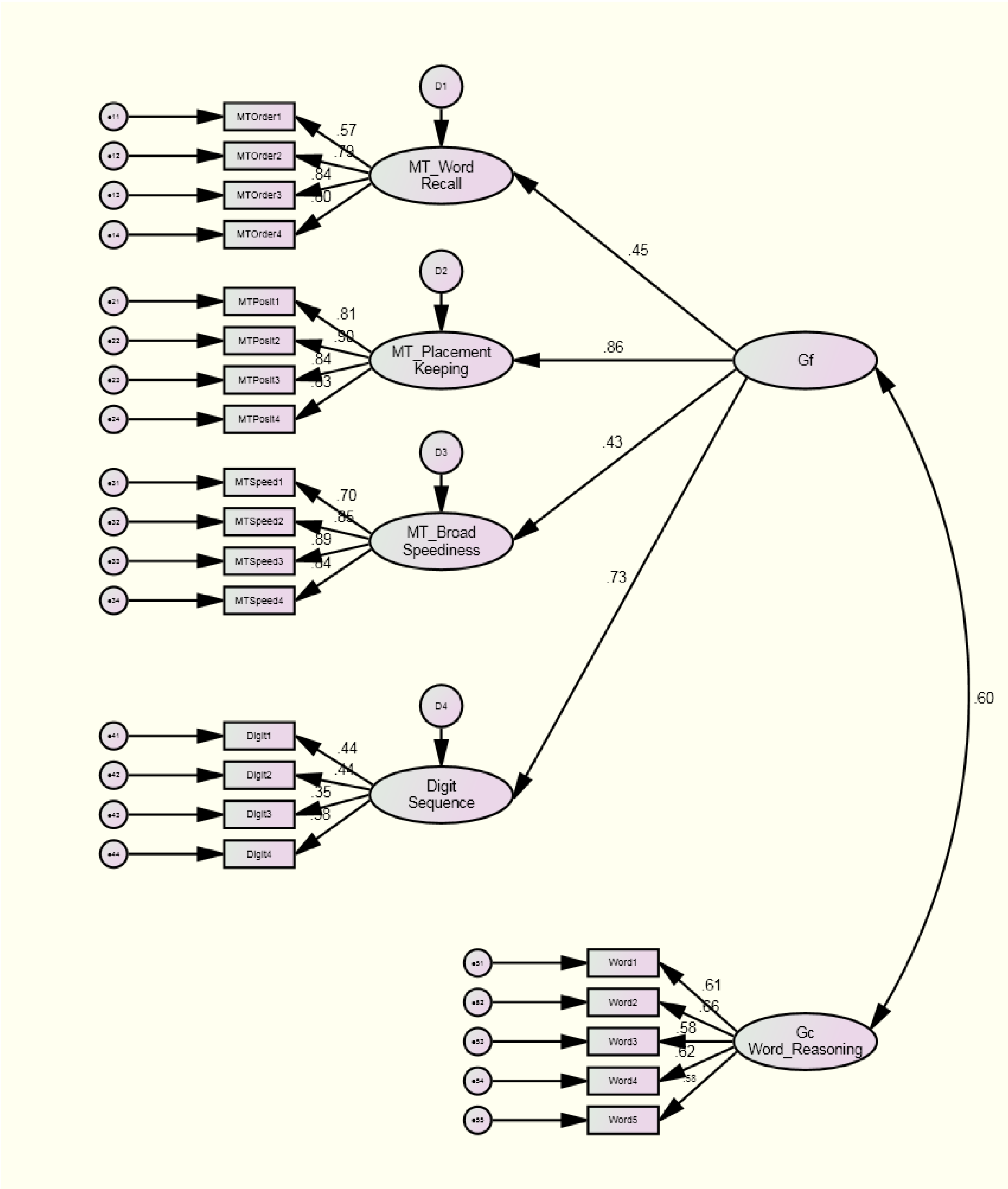


Figure 1.9. Full Multi-Tasks measurement model including single tests.

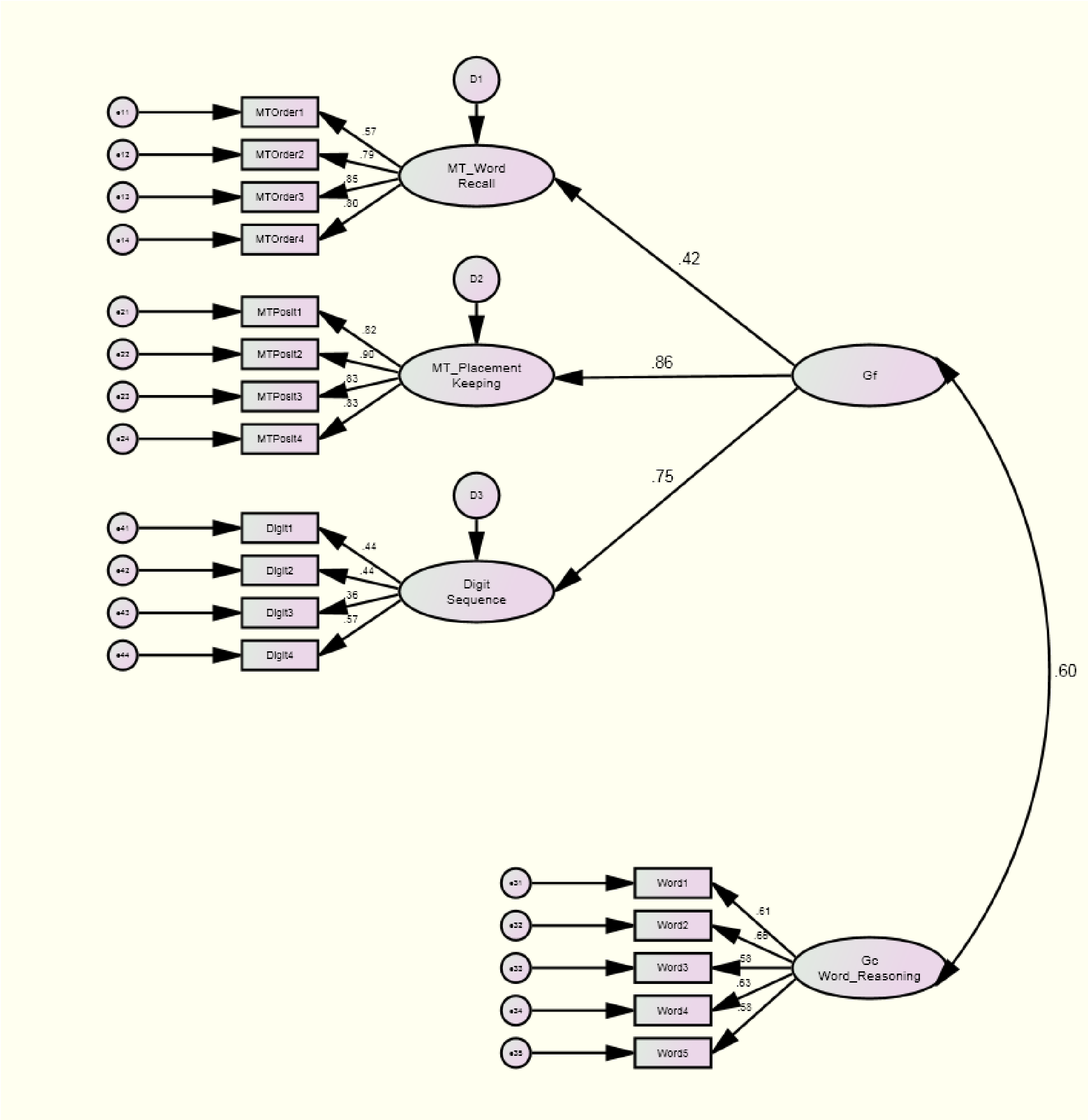


Figure 1.10. Revised measurement model for intelligence.

2.11.4 Predictive Validity of the Intelligence Measures.

To determine the predictive validity of the Multi-Tasks test vis-à-vis the Digit Sequence and Word Reasoning tests, a structural model with the four tests and the outcome measures was tested. This structural model (with the standardized coefficients) is depicted in Figure 1.11 while the fit indices are summarized in Table 1.17. The path coefficients from the constructs to the outcome measures are presented in Table 1.18. The squared multiple correlations for the structural model constructs are given in Table 1.19. The model fit was adequate. Although the SRMR was above the acceptable criterion of .10, the RMSEA value was within the reasonable range and the CFI was close to the acceptable criterion of .95. The ratio of the χ^2 value relative to the degrees of freedom was low and indicates that the model had adequate fit. All parcels, except the third Digit Sequence parcel, loaded on their respective first-order constructs. In addition, the two Multi-Tasks subscales loaded on the second-order Multi-Tasks Test construct.

The Multi-Tasks Test strongly predicted promotional potential ($B = .30$, $p <.001$) and complex problems ($B = .37$, $p <.001$). It did not predict interpersonal relationships. Digit Sequence predicted promotional potential ($B = -.21$, $p <.05$) and complex problems ($B = -.28$, $p <.05$). The negative correlations between Digit Sequence and the outcome measures (in comparison to the positive zero-order correlations in Table 1.11 suggest that Digit Sequence was acting as a suppressor variable in the model. Word Reasoning predicted promotional potential ($B = .36$, $p <.01$), interpersonal relationships ($B = .69$, $p <.001$) and complex problems ($B = .27$, $p = .05$). Thus, it appears that the Multi-Tasks Test had adequate predictive validity and was the strongest predictor of promotional potential and complex job performance.

Table 1.17

Fit indices for the predictive validity model

Index	Value
Chi-square	226.60
Degrees of freedom	162.00
Sig.	.00
Chi-square/df	1.40
Comparative fit index (CFI)	.93
Root mean squared error (RMSEA)	.07
Lower bound 90 % CI	.05
Upper bound of 90% CI	.09
Standardized root mean residual (SRMR)	.14

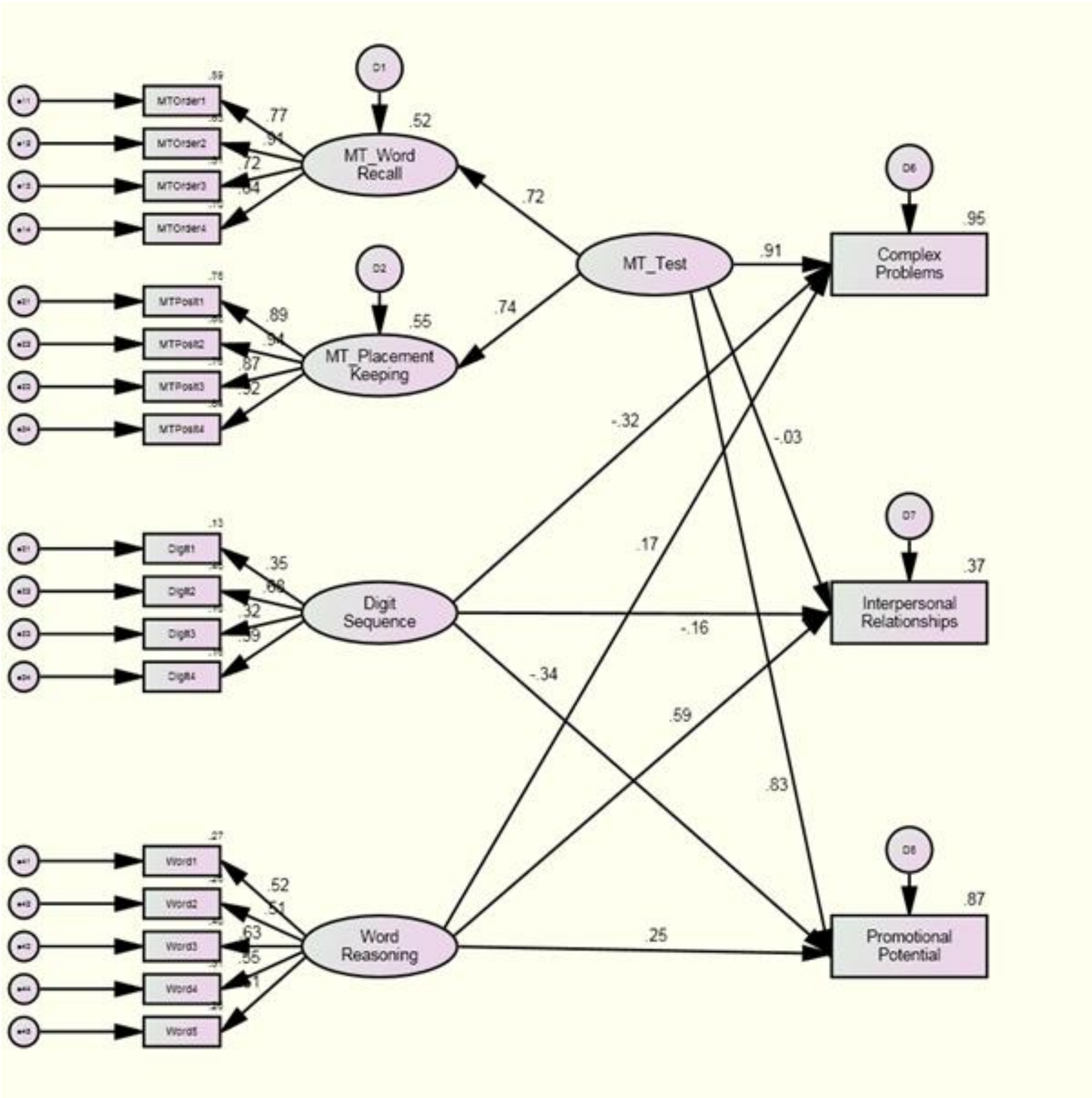


Figure 1.11. Predictive validity model for Multi-Tasks test.

Table 1.18

Path coefficients for the structural model

Relationship	B	SE	β	CR	Sig.
Multi-Tasks to:					
Complex problems	.37	.05	.91	7.56	.000
Interpersonal relationships	-.01	.03	-.03	-.32	.746
Promotional potential	.30	.04	.83	7.26	.000
Digit sequence to:					
Complex problems	-.23	.10	-.32	-2.32	.021
Interpersonal relationships	-.08	.07	-.16	-1.19	.236
Promotional potential	-.24	.09	-.34	-2.31	.021
Word reasoning to:					
Complex problems	.27	.14	.17	1.96	.050
Interpersonal relationships	.69	.19	.59	3.64	.000
Promotional potential	.36	.14	.25	2.58	.010

Table 1.19

Squared multiple correlations for the endogenous constructs of the structural model

Construct	r^2
Multi-tasks word recall	.72
Multi-tasks placement keeping	.74
Complex problems	.95
Interpersonal relationships	.37
Promotional potential	.87

2.12. Discussion

The broad aim of the current two-part study was to explore a new measure of cognitive ability, the Multi-Tasks Test, which employs a competing tasks methodology. Specific aims of the study were to: (1) determine the general properties, including the reliabilities and means of the new test ; (2) assess the predictive validity of Multi-Tasks; (3) determine whether the test measures the broad factors described by the CHC Theory; and (4) investigate the relationship between competing task performance and job performance. An additional aim was to assess differences amongst various demographic groups in performance on the test. The study extended previous research by comparing performance on the Multi-Tasks Test not only with performance on other measures of intellectual functioning, but also with job performance measures of currently employed individuals within Australian organisations.

In relation to underlying structure, factor analysis results suggested that the Multi-Tasks Test can be described in terms of three factors, specifically, Word Recall, Placement Keeping and Broad Speediness. Each of other two tests of cognitive ability (i.e., Digit Sequence and Word Reasoning), used in the study as criterion measures, was found to consist of two types of factors, corresponding to a) number of items correct, and b) speed of response. Structural equation model results found that the Multi-Tasks test was more highly related to Digit Sequence than to Word Reasoning, suggesting that this test may be measuring fluid intelligence more so than crystallized intelligence. This finding supports previous research into competing tasks and the components of intelligence.

In comparing the Multi-Tasks test with the criterion tests, it was found that the new test was more reliable than the others. While this is a promising outcome, the reasonably low reliability of the Digit Sequence test was unexpected as the publisher's manual provided higher reliability coefficients (Omnibus Screening Protocol, 2001). Multi-Tasks was also better at predicting job performance. Although all of the cognitive ability tests were found to be related to at least two of the job performance measures, the Multi-Tasks test was the strongest predictor of two out of the three measures of job performance (i.e., Complex Problem Performance and Promotional Potential). The third job performance measure (i.e., Interpersonal Performance) was predicted by both the Digit Sequence and Word Reasoning tests but not by the Multi-Tasks test.

Word Reasoning was also a strong predictor of job performance. This task reflects an individual's ability to understand, use and reason with language. The ability to understand language and use it to communicate effectively is critical in a workplace. Report writing, emailing, presenting and participating in meetings and informal discussions (to name just a few relevant activities) are common to many different job roles. It is difficult to conceive of any job where some degree of communication is not required. Verbal ability has been shown to be an important predictor of job performance in a range of studies, with a moderate to high operational validity (Bertua, et al., 2005; Gottfredson, 1997; Hunter, 1983 as cited in Schmidt & Hunter, 2004) and has been shown to correlate strongly with general intelligence (Gottfredson, 1997). The particularly strong relationship between Word Reasoning and Interpersonal Relationships in the current study confirms the importance of language for engaging with others in the workplace.

Finally, group differences analyses indicated a significant effect for level of education in Part A (the pilot study) and for gender and age in Part B. In relation to education, in the pilot study those with higher levels of education had higher scores on the Placement Keeping component of the test, and in Part B age the older applicants (who were likely to be applying for more senior positions) had higher scores than the younger applicants. As the older applicants tended to be management applicants, this difference may be due to a higher calibre of candidate applying for these roles. The job demands of management roles and the problem solving complexity of these jobs typically outweigh the demands inherent in lesser, more junior roles. The results of the current study into the effectiveness of the Multi-Tasks test have indicated that there are extensive opportunities to expand the scope of this research. Building on the work of researchers such as Stankov, Fogarty and colleagues who demonstrated the utility of the competing task methodology within non-work settings, and research indicating that Gf measures that tap WM are likely to be good predictors of job performance, the current study has shown a positive relationship between Multi-Tasks, a novel and user-friendly competing tasks measure, and job performance within naturalistic work settings for a sample of Australian job applicants. It was shown to be a more reliable measure than other widely used cognitive ability tests. These findings suggest that the competing task paradigm may have the ability to predict intelligence and job performance over and above single instances of psychometric tests. Furthermore, a strong performance of this test on various statistical measures indicates that the Multi-Tasks test warrants further investigation.

To help establish the Multi-Tasks Test as a valid occupational assessment, it is important to demonstrate that this test not only performs as well as, if not better than, previously used tests of cognitive ability, but that it performs over and above other commonly used assessments attempting to predict job performance. Thus, it is imperative that future studies using the Multi-Tasks test include a well-established personality measure, as well as measures of other known predictors. Being a new measure, it is also important to examine the relationship between Multi-Tasks and job performance in a range of roles, industries and cross-culturally within naturalistic environments. Study 2 will expand upon on study 1 by including a different group of job candidates, alternative measures of job performance and will compare Multi-Tasks to personality in predicting job performance. Research supporting a relationship between one aspect of personality and the Multi-Tasks test will also be outlined. Study 3 will provide a cross-cultural analysis of the factor structure of Multi-Tasks, its reliability and predictive validity in terms of job performance.

2.13. Conclusion

The Multitasks test is a reliable cognitive ability measure which appears to be strongly predictive of job performance. In terms of its place within the CHC Theory of Cognitive Abilities it is most strongly correlated with Gf, indicating that it is a measure of abstract non-verbal reasoning and is likely to be relatively unaffected by education/culture. Its reliability and predictive validity is superior to that of another Gf measure employed in the study (Digit Sequence), which is commonly used in job selection. It is a complex test and research shows that it is highly correlated with general mental ability, making it particularly useful for complex jobs and management roles. Utilising Flash Movie technology, it is also more engaging and interesting than many traditional measures. Further, it is an efficient measure, allowing for reliable intelligence data to be collected in a reasonably short time-frame. As it is a relatively new test further research is required, particularly with a variety of job types and cultural groups. Study 2 and 3 in this series will further investigate the reliability of the measure, how it compares to other measures used in personnel selection, and the ability of Multi-Tasks to predict job performance in different settings, with people of different cultural backgrounds and in another country.

CHAPTER 3

STUDY 2

COMPETING TASKS AS MEASURES OF INTELLIGENCE AND PREDICTORS OF JOB PERFORMANCE

3.1 Competing Task Research

Competing task research has re-emerged as technological advances have facilitated the cost-effective construction and administration of these tasks. The competing task paradigm may be particularly useful in personnel selection as it represents a practical means for manipulating complexity in a novel manner, without the need for an extensive battery of intelligence tests; providing an indication of the extent to which an applicant will be able to cope with a range of workplace demands. Using this paradigm, two psychometric tests assessing cognitive abilities are presented concurrently, meaning that the task can be completed in a relatively short period of time, and provide more information than single tests. They are also more novel than standard cognitive tests and have the potential to engage the participant as they are visually stimulating. Indeed, earlier studies (e.g., Fogarty & Stankov, 1982; Stankov et al., 1989) have supported this proposition.

The findings of Study 1 are a promising indication of the potential value of the competing tasks measure, Multi-Tasks test, in personnel selection. Being a new measure, it is important to investigate how it fits within the broad domain of intelligence and how it compares with other well-known measures used in personnel selection. This study will expand on Study 1 to examine the relationship between the Multi-Tasks test and a wider range of cognitive measures, in an effort to better define the test's place within the CHC Theory of Cognitive Abilities (McGrew, 2004; 2009) and compare it with a range of other intelligence tests. Additionally, it will be examined within the context of one of the most popular forms of employment test, the personality inventory (Barrick, Mount & Judge; 2001; Hogan, 2005; Hogan & Holland, 2003).

In Study 1 the Multi-Tasks test and the other cognitive ability measures included a component that measured the speed of response ('speediness'). These speediness components operated differently, in terms of their correlations with Multi-Tasks and their factor loadings on crystallized intelligence (Gc) and fluid intelligence (Gf). For instance, the Word Reasoning (Gc task) speediness was positively correlated with one of the Multi-Tasks measures, while Digit Sequence (Gf measure) speediness was not correlated with either of them. The factor loading of the Multi-Tasks speediness on to Gf was low and a model incorporating the broad speediness components of Word Reasoning and Digit Sequence had inadequate fit. A revised model that was a better fit for the data was created by dropping all measures of speediness. This is partly explained by the literature that indicates the role of processing speed in intelligence is complicated and controversial. For instance, processing speed has been shown to be important in performance on tests of competing tasks and general mental ability (McGrew, 2004; Stankov & Roberts, 1997). However, there is also research which suggests that its importance as a factor underlying general mental ability has been overstated and its relationship to competing tasks and intelligence appears to be an indirect one (McGrew, 2004; Stankov & Roberts, 1997). Introducing a speed component may affect the psychometric properties of the test and speeded and unspeeded tests do not necessarily tap the same constructs (Wilhelm & Schulze, 2002). Further, speed is likely to be predictive of performance in job roles where working quickly is of utmost importance (e.g., factory production line work), not in professional jobs, such as those which the participants in Study 1 and the current study were employed. For these reasons, processing speed will not be considered in Study 2.

3.2 Multi-Tasks, Gsm and Working Memory

An important component of the CHC theory not included in Study 1 is Gsm, a limited capacity system which allows one to maintain awareness of events that have just occurred (e.g., the last minute or so). Information in the system is lost quickly unless cognitive resources are used to maintain it in immediate awareness (e.g., repeating a telephone number over and over in order to remember it) (McGrew, 2009). Gsm is measured through tests of immediate memory span (Stankov, 1988).

There are claims in the literature that, together with Gf, Gsm plays an important role, and a greater role than the other CHC components, in competing task performance (e.g., Stankov, 1988; however, see Stankov & Myors, 1990).

An important process underlying this relationship is the role of working memory (WM). Research has shown a relationship between intelligence (particularly Gf), WM and performance on competing tasks (Conway et al., 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, & Conway, 2005). WM is thought to be composed of a short-term memory store and an executive component which manages attention and allocates resources to refreshing and working on the content in the memory store, particularly in the presence of interference (Kane & Engle, 2002; McGrew, 2004; Stankov, 1999). König, Bühner and Mürling (2005) found that WM, attention and Gf emerged as the most important predictors of performance in the competing task paradigm, with working memory resulting in the highest correlations with performance. They surmised that because of the complexity of the competing task paradigm it was the higher level cognitive processes, specifically executive control of attention (i.e., working memory) that accounted for most of the individual differences in performance (König et al., 2005).

Gsm represents the ‘passive’ component or short-term memory store component of WM (Engle et al., 1999; McGrew, 2004; Stankov, 1988). Using memory span (Gsm) and WM measures from the Woodcock Johnson III Tests of Cognitive Abilities (McGrew & Woodcock, 2001). McGrew (2004) found latent factor correlations between the measures of .67 (6-8 years of age), .79 (9-13 years), .82 (14-19 years), .84 (20-29 years), and .80 (40-90+ years). He concluded that there is a strong correlation between Gsm and working memory, which increases with age. The correlation is not perfect however, as memory span is not the same as, but rather a component of, WM. Gsm has been shown to have a higher correlation with competing than single task performance (Stankov, 1988).

3.3 Gf, Gsm and Task Complexity

It has been suggested that as tasks increase in complexity, the WM load of the task increases, as does the correlation of the task with Gf (Spilsbury et al., 1990). Evidence to support the mediating role of complexity in the relationship between WM and Gf can be found in the intelligence literature. For example, Carpenter, Just and Shell (1990) found that a commonly used test of Gf, the Raven's Standard Progressive Matrices test (Raven et al., 1998) measures abilities which are thought to be associated with WM load, such as the generation and maintenance of goals. This test was also found to be highly complex (Carpenter et al., 1990).

Roberts, Beh, and Stankov (1988) and Roberts, Beh, Spilsbury and Stankov (1991) showed that correlations between Gf and competing tasks increased as task complexity increased (extra bits of information were added), while correlations of performance with Gc were mostly non-significant. However they did not find strong correlations between Gsm and competing task performance. This finding is contrary to Stankov's (1988) study where Gsm had a higher correlation with competing than single task performance. This makes intuitive sense, because Gsm and WM are related concepts and strongly correlated. It may be that the different research outcomes result from different conceptions and measures of WM used.

As hypothesised in Study 1 Gf demonstrated a stronger relationship with performance on the Multi-Tasks test than Gc. The effects of Gsm were not partialled out however. The direct relationships between Multi-Tasks, Gf and Gsm and the effect on the relationship between Multi-Tasks and Gf when Gsm is introduced to the analysis is of interest given the previous research showing the relationship between WM, Gsm, and competing tasks.

Having outlined how the Multi-Tasks test relates to the CHC Theory of Cognitive Abilities the discussion now turns to the other main variable of interest in this study, personality assessment. Personality is of interest because it may play a role in predicting performance on cognitive tasks relevant to the Multi-Tasks test and because it is a popular and widely used measure in personnel selection. While its use is pervasive and much of the research about personality assessment is positive, there is also research to suggest that not all aspects of personality are reliable predictors of job performance in all industries (Barrick et al., 2001).

3.4 Personality Assessment

Personality assessment is a broad term that includes interpretations of scores on personality measures and behaviour in the context of the culture of the individual being evaluated. As a scientific endeavour, the purpose of personality assessment is to determine and measure individual differences that are indicative of behaviour, cognition and attitudes (Oze & R & Reise, 1994; Hogan, 2005).

The Five Factor Model of personality (FFM; Digman, 1990; Goldberg, 1993) will be the theoretical basis for the exploration of personality in this study. The FFM comprises five personality components: extraversion, agreeableness, conscientiousness, openness to experience, and emotional stability. The empirical evidence supporting this model and its relationship with job performance will be discussed in the sections that follow.

Of most relevance to the role of personality in performance on cognitive tests is extraversion. Extraversion includes traits such as sociability, assertiveness and ambition.

3.5 Personality and Cognitive Abilities

Research has shown an interesting set of relationships between extraversion and cognition, particularly working memory which, as previously discussed, is an important factor to consider in relation to the Multi-Tasks test. Extraversion is largely about the way people interact with and relate to others. Extraverts tend to seek out social interaction, enjoy the company of others and to be assertive in their interactions. Introverts are more likely to be shy and reserved and may feel uncomfortable around other people, preferring to be alone or in small groups.

Eysenck's biological theory of personality explains these differences in terms of the amount of neuronal activity (or 'arousal') in the brain of extraverts and introverts. Dopamine cells in the ascending reticular activating system (ARAS), a structure which extends from the brain stem to the cortex (Eysenck & Eysenck, 1985; Rammsayer, 1998) modulate the amount of activation in the brain, and also play a role in tasks requiring a motor response. Studies have shown that introverts and extraverts respond differently to pharmacologically induced changes to dopamine cells, with introverts being more susceptible to these changes (Le Moal & Simon, 1991; Rammsayer, Netter & Vogel, 1993). PET brain scans have shown that introverts have higher dopamanergic activity in certain parts of the brain associated with the ARAS than extraverts (Fischer, Wik, & Fredrikson, 1997).

The outcomes of these studies suggest that introverts have a higher level of natural 'arousal' and can therefore be more easily over-stimulated by sensory input, for instance interacting socially with others. Alternatively, extraverts have lower levels of arousal and thus seek out external stimulation.

3.5.1 Extraversion and WM.

Performance on WM tasks has also been shown to be related to extraversion. Nuclei in the ARAS project to the dorsolateral prefrontal cortex (DLPFC), which is involved in WM and the basal ganglia which modulates the release of dopamine (Lieberman, 2000; Martin, 1996). Both the DLPFC and the basal ganglia rely on dopamine for their normal functioning (Lieberman, 2000). It is therefore likely that extraverts and introverts who differ in their arousal via the ARAS are likely to perform differently on working memory tasks.

This hypothesis has been indirectly investigated in previous studies; however these have been limited by the empirical understanding of working memory (WM) at the time of the studies. For instance, Howarth (1963) and Tanwar and Malhotra (1992) (as cited in Lieberman, 2000) used a digit span task (repetition of increasingly longer series of numbers) to examine the memory span of introverts and extraverts, with the former finding no significant difference and the latter finding significantly longer spans in extraverts. As previously discussed, memory span is the short-term memory store component of WM. It relies very little on the executive component (or central executive) which is thought to be located in the DLPFC and is responsible for monitoring the contents of the memory store and comparing it with external stimuli (Lieberman, 2000). The short-term storage systems that hold the WM components are located elsewhere in the brain and are not associated with the ARAS. It is therefore likely that introverts and extraverts will differ more strongly on tasks that draw on the central executive component of WM (Lieberman, 2000).

A number of studies have investigated a more comprehensive and complex measure of WM to investigate the hypothesis. Howarth (1969, as cited in Lieberman, 2000) used a response competition paradigm with a paired-associates list learning task. In this task participants learned pairs of words and then in a recall task one of each pair, i.e. list A was used as a retrieval cue for the other, list B. They then manipulated the task so that after learning the first set of associations, the pairs were changed so that the words from list A were now paired with different words from list B. Extraverts required fewer learning trials than introverts on this task, demonstrating that they were better able to inhibit prior over learned responses that were errors in the second part of the task, an ability drawing on the central executive component of WM (Lieberman, 2000). Lieberman (2000) used a memory scanning task involving a memory set of 1 to 6 digits to be held in mind during a block of trials. During the trials subjects had to respond as quickly as possible to indicate whether the number presented was a member of the memory set. Performance required holding the numbers in mind (Gsm), and attending to the trial stimulus and comparing it with the numbers held in mind (central executive component of WM). Regression analyses indicated that extraverts had faster reaction times than introverts, thus were able to compare the contents of the information held in the short-term memory store to the stimulus being presented and make a decision more quickly than introverts (Lieberman, 2000).

Lieberman (2000) explained these results in terms of ‘cognitive busyness’. Introverts are naturally more cognitively ‘busy’ than extraverts. Their more active ARAS leads to a greater production of dopamine and thus higher levels in the DLFPC (Fischer et al., 1997). This is a problem because there appears to be a narrow range of dopamine that is optimal, beyond which performance on WM tasks becomes adversely affected (Arnsten & Goldman-Rakic, 1998). Further, when a higher degree of complexity is introduced (e.g. a response inhibition task) extraverts appear better able to deal with the increased complexity than introverts, who are already more naturally stimulated that extraverts.

3.5.2 Implications for Performance on Multi-Tasks and job performance.

These studies suggest that extraversion may influence performance on the Multi-Tasks tests, which is complex and draws heavily on WM; thereby indirectly impacting job performance. The research indicates that due to their lower level of natural cognitive arousal or ‘busyness’ and therefore greater ability to deal with introduced cognitive complexity, extraverts perform better on tasks that draw on the executive component of WM, which is modulated by the same part of the brain responsible for the level of cognitive arousal. Numerous previous studies have demonstrated that competing tasks measures are complex and draw on the executive component of WM (e.g., König, et al., 2005; Stankov, 1988) and it is therefore hypothesised that there will be a relationship between Multi-Tasks performance and extraversion. There is substantial evidence that shows a direct effect of extraversion on job performance.

In terms of personality more broadly personality assessment has played an important role in personnel selection for at least the last 20 years and there is a large body of literature to support a relationship between personality and job performance. The history and current state of personality assessment in job performance literature, particularly the FFM will be outlined in the sections that follow.

3.6 Personality Assessment and Job Performance

There are two main reasons for considering the relationship between personality attributes and job performance. Firstly, personality assessment has gained considerable popularity and support in research and practice over the last 20 years, and is now widely considered to be a valid indicator of job performance across industries (Barrick & Mount, 1991; Barrick, et al., 2001; Hogan, 2005; Hogan & Holland, 2003). Secondly, consideration of personality will help to further define the relationship between Multi-Tasks and job performance by examining both the direct relationship and its relative importance when compared to other widely used measures. Together with general mental ability, measures of at least some personality variables appear to be generally valid for all jobs (Bartram, 2004). These concepts will be further discussed in the following sections.

3.6.1 Personality Research in the 20th and 21st Centuries.

Since the development of the Big Five model, personality in the workplace has been frequently examined. The development of increasingly sophisticated meta-analytic techniques and general Web-based technological developments are considered to have radically changed the way in which personality assessment is viewed in the workplace (Goodstein & Lanyon, 1999; Walsh & Eggerth, 2005). Personality measures can supplement other measures such as ability tests and add significant incremental validity to a selection battery (Bartram, 2004). Further, when ability and personality measures are combined, group differences often seen in tests of ability (e.g., cultural group differences) on the composite score measures tend to be reduced, thus minimising bias (Bartram, 2004; Schmidt & Hunter, 1998; Walsh & Eggerth, 2005). In addition, personality assessment does not require prior job experience and, therefore, may be implemented in any level of the organisation. Personality can also be measured using both self and observer ratings, including multiple raters, thereby increasing the reliability and validity of assessments (Oh, Wang, & Mount, 2010; Zimmerman, Triana, & Barrick, 2010). A further reason for this increased enthusiasm is attributed to advances in tests and techniques associated with recruitment and selection (Bartram, 2004).

3.6.2 The Five Factor Model of Personality.

Based on the trait model of personality, the Five-Factor model (FFM) (Digman, 1990; Goldberg, 1993) has received broad recognition and acceptance due to the extensive empirical evidence supporting the model and the predictive capabilities of its factors. Overriding Eysenck’s three factor model of personality, the FFM provides an effective framework for summarising the relationships between personality traits (Oze & Reise, 1994). As such, it is a widely accepted model and remains dominant in contemporary research (McCrae & Costa, 2008; Oh et al., 2010).

The Five Factor’s ‘Big Five’ are currently labelled: extroversion (E), agreeableness (A), conscientiousness (C), openness to experience (O), and emotional stability (ES). These five factors exist on a bipolar scale; that is, there is a ‘positive’ and ‘negative’ pole of each dimension, neither of which represent nor indicate pathology (Walsh & Eggerth, 2005). Descriptions of the Big Five are given in Table 2.1 and are based on Barrick and Mount (1991, pp. 3 - 5) and Barrick et al. (2001, p. 11).

Table 2.1 Descriptions of Big Five factors

Factor	Associated traits
Extraversion	Being sociable, dominant, gregarious, talkative, assertive, adventurous, excitement-seeking, active, energetic, positive, and ambitious
Agreeableness	Being courteous, good natured, flexible, trusting, cooperative, compliant, forgiving, empathic, caring, soft hearted, and tolerant
Conscientiousness	Being careful, thorough, responsible, dependable, organised, efficient, persevering, hardworking, and achievement oriented
Emotional Stability	being calm, composed, poised, resilient, adaptable, and self-reliant. The negative pole of this construct, neuroticism (N), is defined by anxiety, hostility, depression, and personal insecurity
Openness to Experience	being imaginative, unconventional, artistically sensitive, intellectual, curious, creative, polished, broad-minded, original, and independent

The Big Five have been empirically established by factor-analytic research and the traits are considered to remain stable over time (Schmidt & Hunter, 2004), with evidence indicating substantial heritability (e.g., Jang, Livesley, & Vemon, 1996). The model has been researched in various languages and has been found to apply to range of cultures (Schmidt & Hunter, 2004). It also generalises across different sources of ratings and measures (Judge & Ilies, 2002).

Since the development of the Big Five model, numerous studies have found significant relationships between the five personality factors and measures of organisational behaviours, such as job performance, leadership, work attitudes, and motivation (Ones et al., 2007; Walsh E & Eggerth, 2005). This review will address the relationship between the Big Five personality factors and job performance, as well as facets of job performance and other performance-related criteria. By adapting the FFM taxonomy, researchers have been able to develop specific hypotheses about the relationships between personality factors and performance-related outcomes. A review of the research on the relationship between the FFM and job performance follows, commencing with the early meta-analytic research, and leading into a discussion of each of the factors individually.

3.6.3 Early Meta-Analytic Evidence for the Relationship between the FFM and job performance.

Several early meta-analyses and investigations of meta-analytic data provided sound support for the relationship between the five personality factors and job performance. The first large meta-analysis was conducted by Barrick and Mount in 1991. Its purpose was to investigate the role of the Big Five in predicting workplace behaviour. This study revealed a significant relationship between the Big Five personality dimensions and several performance indicators. More specifically, O and E were found to predict training proficiency, and E was found to be a significant predictor of performance in occupations involving social interaction.

McNeely and Meglino (1994) found positive relationships between C, A and extra-role behaviours. Barrick, Stewart, Neubert and Mount (1998) reported that E and ES predicted performance at the team level. A meta-analytic review conducted by Hurtz and Donovan (2000) determined two of the Big Five factors to predict job performance, namely, C (mean correlation of .20) and ES (mean correlation of .13).

The next major study, Barrick et al. (2001), supported and extended previous meta-analytic research. C emerged as the strongest predictor of overall performance across all investigated occupations. ES was also found to be a generalisable predictor of overall job performance; however, its relationship to specific performance criteria and occupations was less impressive than that of C. The other three Big Five factors, E, O, and A, were not directly predictive of overall job performance, however, these factors were found to predict teamwork and training.

The next section will explore the mechanisms underlying the relationships between each facet of the FFM personality model and job performance.

3.6.4 Relationship between each of the Five Factors and Job Performance.

3.6.4.1 Conscientiousness.

The C factor of the Big Five is regarded as the strongest and most generalisable predictor across job performance measures in all occupations studied (Barrick & Mount, 1991; Barrick et al., 2001; Schmidt & Hunter, 1998). This is largely because its characteristics (i.e., persistence, planning, responsibility etc.) are considered to be important attributes for satisfactory job performance in any occupation (Barrick & Mount, 1991). In general terms, highly conscientious individuals tend to be more motivated to achieve better performance on the job (Judge & Ilies, 2002). This is thought to be due to their tendency to plan, set goals, and persist at tasks (Barrick & Mount, 1991; Hurtz & Donovan, 2000). Indeed, recent meta-analyses have supported this conclusion (Ones et al., 2007).

The relationship between job performance and C is particularly notable when job (task) complexity acts as a moderating variable. Most researchers agree that higher complexity jobs require higher levels of C (e.g., Le et al., 2011; Mount, Oh & Burns, 2008). However, research also shows that various levels of C may be suited to different tasks (Mount, et al., 2008). For example, in low complexity jobs that require speed, the deliberate and cautious traits of high C are unlikely to benefit performance. In contrast, highly complex jobs often require accuracy, creativity, and persistence, which are features of high C. Alternatively, the research that supports the relationship between C and performance, moderated by task complexity, proposes a curvilinear relationship, as opposed to a linear one (Le et al.). Similar to the relationship proposed for ES and performance, it is suggested that an ‘optimal’ level of C exists. Beyond this level, a potential detriment to performance may be evident, for example, an individual being overly cautious and deliberate (Le et al.).

As it consistently predicts job performance across occupations, C is considered to function in a similar manner to general mental ability (GMA), or general intelligence

(Schmidt & Hunter, 2004; Zimmerman et al., 2010). As demonstrated by Mount, Barrick, and Stewart's (1998) meta-analysis, this is particularly evident when C and its components (achievement and dedication) are conceptually related to job criteria. That is, components of C are considered to predict the specific measures of job performance more accurately than the global measures of job performance (Mount & Barrick, 1998; Walsh & Eggerth, 2005). Performance-related outcomes will be addressed in the next section.

3.6.4.2 Emotional Stability.

ES has been established as a potentially generalisable predictor of overall job performance, though it achieves a lower mean correlation than C, p = .13 (Barrick et al., 2001). It is intuitively plausible that individuals who are most likely to perform well are calm, adaptable, self-reliant and resilient, and conversely, that neuroticism (i.e., nervousness and volatility) is unlikely to facilitate satisfactory performance (Barrick & Mount, 1991). However, its relationship to different occupations and more specific performance criteria is less consistent. Some researchers have argued that the relationship between performance and emotional stability may be curvilinear, as opposed to strictly linear (Le et al., 2011; Ones et al., 2007). That is, the predictive validity of ES is unlikely to increase (and may decrease) with higher scores because, beyond a 'sufficient' level, there may be no benefit from very high scores (Le et al., 2011). Nevertheless, ES appears to be related to overall performance in skilled and semi-skilled occupations, including police (p = .11), and customer service (p = .12) (Ones et al., 2007).

3.6.4.3 Agreeableness, Openness to Experience and Extraversion.

According to Barrick et al.'s (2001) study, A, O, and E do not predict overall job performance; however there is a relationship between these factors and specific types of job performance. These factors do appear to be predictive of performance in jobs involving frequent interaction or cooperation with others. For example, E is a useful predictor of performance in roles that involve leading, mentoring or persuading others, skills that require sociability, assertiveness and ambition (e.g., managerial, sales and police occupations) (Barrick et al., 2001). A is more a function of interpersonal interaction when the interaction involves cooperation, helping and nurturing, and it is therefore considered to facilitate functioning in a team environment (Barrick et al., 2001).

Openness to experience is considered to measure an individual's attitude towards learning, and is, thus, often associated with one's training proficiency (Barrick& Mount, 1991). It is considered that individuals who exhibit the traits of openness to experience are more likely to be willing to undertake and benefit from training endeavours (Barrick et al., 2001).

The most notable effect sizes found in Barrick's et al. (2001) study included that of O and training performance and A and teamwork performance, both with values over .30. Ones et al. (2007) found that A predicted customer service ($p = .17$) and police ($p = .10$) occupations; while O and E predicted customer service ($p = .15$ and $p = .11$, respectively).

3.7 Implications and Methodological Concerns

The most robust finding in the relationship between personality and job performance is that C predicts a wide range of job performance criteria in a wide range of occupations. Some support is also evident for ES, although findings are weaker and less consistent. Finally, the remaining Big Five factors, while not predictive of overall performance criteria, are found to be associated with facets of job performance.

From a practical sense, it must be noted that predictive validity is the most important contribution of assessments (Schmidt & Hunter, 1998). While often statistically significant and valuable to the body of research, the predictive power of the Big Five factors remains comparatively low to that of other assessment methods. In predicting overall job performance, the most impressive results do not extend beyond 7.3% of outcome variance

(Ones et al., 2007). However, the predictive validity of personality assessment improves, often substantially, when the outcome comprises alternate performance-related criteria.

Nevertheless, the validity and predictive power of personality assessment has its critics. Indeed, personality assessment is considered to provide less explained variance than GMA (Schmidt & Hunter, 1998), though a growing body of evidence suggests that it may contribute to the overall validity or, at a minimum, extend understanding of behaviour in the workplace.

Indeed, one opinion is that the predictive power of personality assessment does not lag far behind that of GMA. A longitudinal study conducted by Judge et al. (1999) investigated GMA and personality factors at 12 years of age, in relation to occupational outcomes at 41 to 50 years of age. As expected, GMA was highly predictive of adult income and occupational level (correlations of .53 and .47-.71, respectively). However, this research also revealed that three of the Big Five personality factors were found to predict occupational level and income; namely, C (.49 and .41, respectively), O (.32 and .26, respectively), and ES (-.26 and -.34, respectively). It is evident that the correlations produced by the C factor were only marginally smaller than those associated with GMA, and potentially add theoretical understanding to the study of occupational outcomes.

However, other research has shown that personality measures provide little incremental validity over and above GMA (e.g., Le et al., 2011). However, Le et al. (2011) also acknowledged that any incremental validity with personality measures is beneficial, particularly if it is not associated with increases in cost. By the same token, incremental validity may increase the efficiency and effectiveness of decisions, potentially reducing expenses and increasing savings or profits over time (Le et al., 2011; Schmidt & Hunter, 1998). Moreover, in the context of selection, personality assessments are considered to add a relatively gender- and race-neutral increment to the predictive validity of GMA measures (Walsh & Eggerth, 2005).

Research indicating that personality may influence performance on cognitive tasks that tap into WM provides further support for its inclusion into job performance studies, particularly those which include complex measures of cognitive ability.

In conclusion, research has revealed some promising findings regarding personality assessment and the FFM in particular. The purpose of personality assessment is, evidently, not to become an alternative to GMA tests. Rather, there is strong support for utilising personality assessment as a supplementary technique. Research suggests that the Big Five factors may be valuable predictors of job performance, performance-related outcomes, and performance on complex cognitive tasks such as the Multi-Tasks measure, the implications of which suggest that those who are interested in such outcomes may benefit from undertaking personality assessment, provided they have correctly determined the outcome criteria and purpose of the assessment.

3.8 Study Aims

The main aim of the current study was to extend the findings of the first study in this series, which explored a new measure of cognitive ability, the Multi-Tasks test, which employs a competing task paradigm. Specific aims of the study were to: (1) determine if the Multi-Tasks test can be defined as two factor model, similar to that shown in Study 1, but excluding a speed factor; (2) based on an extended protocol of cognitive ability measures, determine if Multi-Tasks measures three of the broad factors described by the CHC Theory of Intelligence, namely fluid Intelligence (Gf), crystallized Intelligence (Gc), and short-term Memory (Gsm); (3) compare the predictive validity of the Multi-Tasks test with that of the single tests measuring Gf, Gc and Gsm in predicting job performance as measured by a four job performance indicators; (4) identify whether personality has an impact on performance on the Multi-Tasks test; (5) compare the predictive validity of the Multi-Tasks test with that of the personality measure in predicting job performance. Further, to explore demographic differences in performance on the Multi-Tasks test, specifically whether performance varies by age, gender and level of education.

It is hypothesised that Gf and Gsm will have the strongest correlations with the Multi-Tasks test, and that while all ability measures will be related to job performance, the Multi-Tasks test will be a more reliable and stronger predictor of job performance. It is further hypothesised that the personality measure will predict job performance, but not as strongly or reliably as the Multi-Tasks measure. In terms of the relationship between personality and the Multi-Tasks test, it is hypothesised that those individuals higher on extraversion will perform better on the Gsm task and the Multi-Tasks test.

Three models were proposed to explore Multi-Tasks and its place within the CHC model of intellectual functioning. Figure 2.1 outlines the proposed model for the underlying structure of the competing tasks measure (Multi-Tasks). Based on the results of Study 1 it was hypothesised that the underlying dimensions of the competing task measure would include factors relating to each of the competing tasks (word recall and placement keeping). As previously discussed, the speed component of the measures is being excluded from this study.

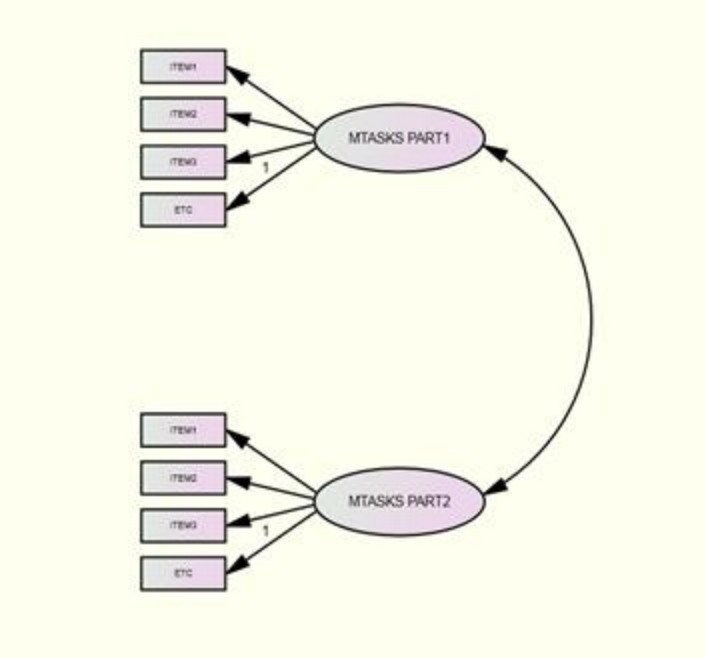


Figure 2.1. Proposed measurement model for Multi-Tasks.

Figures 2.2 and 2.3 outline the hypothesised relationships between the competing tasks measure and other well-known measures used in job selection and the CHC theory of intelligence. Based on the outcomes of Study 1 and previous research (e.g., König, et al., 2005; McGrew, 2009; McGrew & Woodcock, 2001; Stankov, 1988) it was hypothesised that the components of the Multi-Tasks measure would be strongly associated with Gf, as would the Digit Sequence and Matrices tasks, whereas the Word Reasoning and Reading Comprehension tasks would load onto the Gc factor. Digit Sequence and Matrices are novel non-verbal problem solving tasks, while Word Reasoning and Reading Comprehension are verbal language reasoning and comprehension tasks.

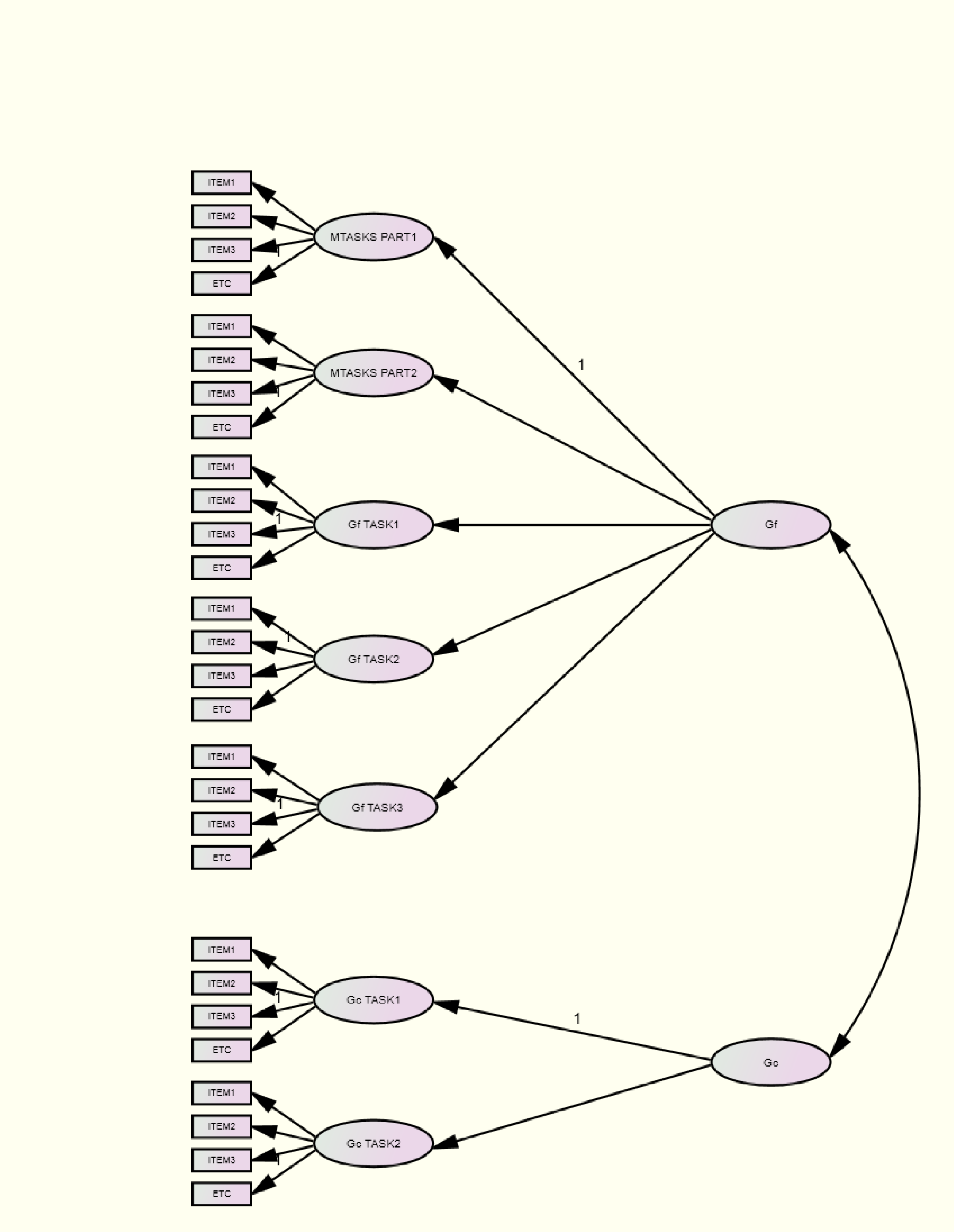


Figure 2.2. Proposed measurement model for Multi-Tasks and other intelligence measures.

Introducing Gsm into the model was proposed to further define the place of Multi-Tasks within the CHC theory. It was hypothesised that because competing tasks have been shown to have a higher correlation with Gsm than single tasks (Stankov, 1988) and due to the high demands of the Multi-Tasks test on WM which is highly correlated with Gsm (McGrew, 2004), Multi-Tasks would have a strong relationship with Gsm. As would Digit Span, which is a Gsm task (McGrew, 2009: McGrew & Woodcock, 2001).Matrices and Digit Sequence would continue to load onto the Gf factor and Word Reasoning and Reading Comprehension on to the Gc factor. This is represented in Figure 2.3. Speed was removed from the model in Figure 2.3 as Study 1 revealed that a model incorporating speed measures did not fit the data well. It was hypothesised that this would also be the case in the current study.

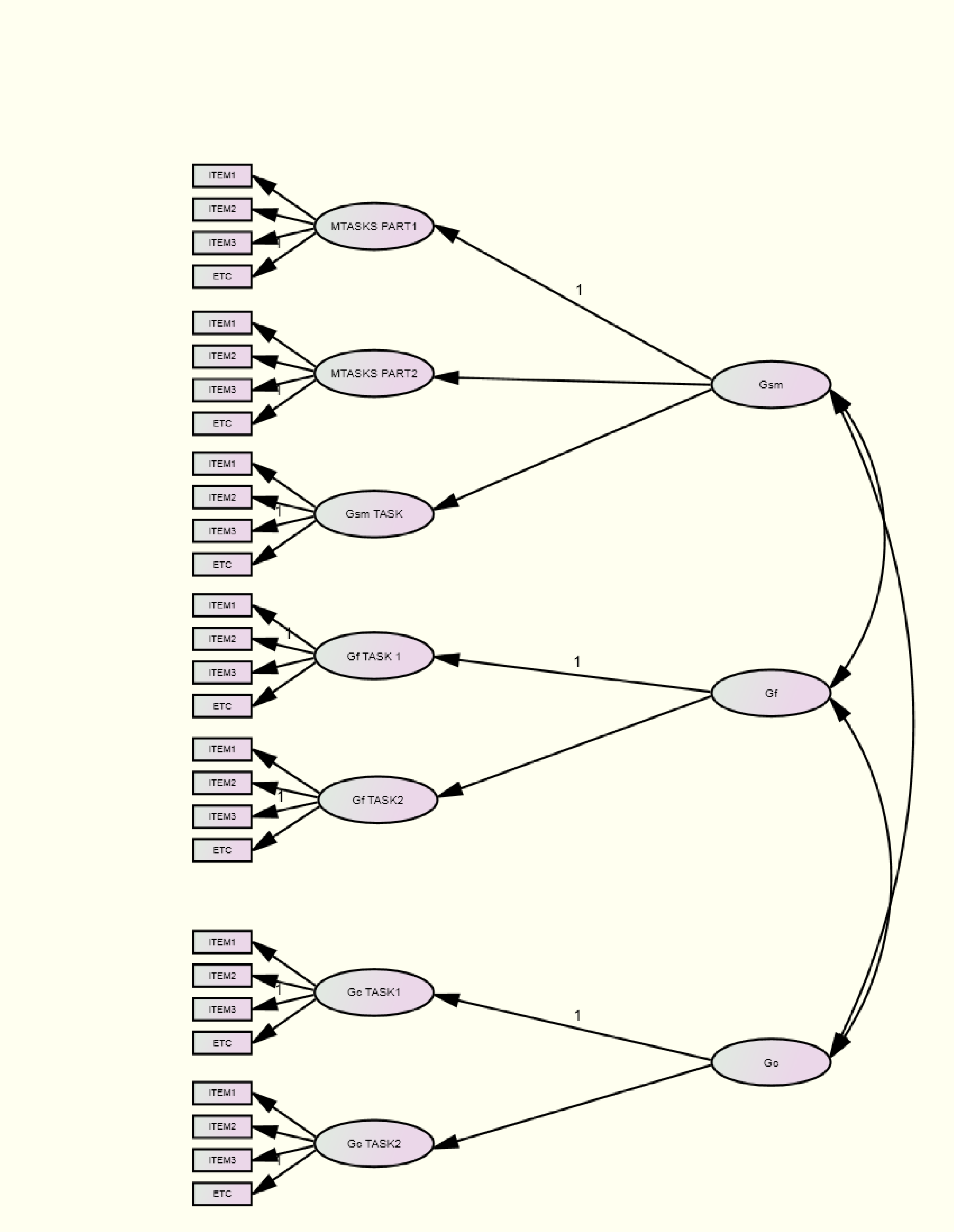


Figure 2.3. Proposed measurement model for Multi-Tasks and other intelligence measures (2).

3.9 Method

3.9.1 Participants.

A total of 1391 job applicants for a range of banking and finance industry positions in different locations within one organisation in Australia participated in the study. The group consisted of 459 (33%) males and 930 (66.9%) females. Two participants did not specify their gender. Age and education level of the participants are shown in Tables 2.2 and 2.3 respectively.

Table 2.2

Age of participants (N =1391)

Variable	Frequency	Percentage
Age		
18 to 24	592	42.6
25 to 29	361	26.0
30 to 34	150	10.8
35 to 39	85	6.1
40 to 44	70	5.0
45 to 49	61	4.4
50 to 54	40	2.9
55 and above	30	2.2
Missing	2	.1

Table 2.3

Education level of participants (N = 1391)

Variable	Frequency	Percentage
Education level		
Secondary certificate	267	19.2
Higher school certificate	458	32.9
Tafe qualified	209	15.0
Took college courses	107	7.7
Undergraduate university degree	230	16.5
Postgraduate university degree	70	5.0
Missing	50	3.6

3.9.2 Materials.

The study employed six measures of cognitive ability, a personality inventory and four measures of job performance. The ability measures included two tests that were used as proposed criterion measures of fluid intelligence (Gf), one of which was used in Study 1 (Digit Sequence) and one which was not (Matrices), and two tests that were used as proposed criterion measures of crystallized intelligence (Gc), again one of which was used in Study 1 (Word Reasoning) and one not (Reading Comprehension). While Digit Sequence and Word Reasoning are widely used measures of cognitive ability, they had moderate reliability coefficients in Study 1. Matrices and Reading Comprehension were therefore included to provide additional measures and to more thoroughly test Gf and Gc. New tests used in the current study were Digit Span, a proposed test of short-term memory (Gsm). The sixth test was the new experimental measure called the Multi-Tasks test, which was also used in Study 1. The measures of job performance consisted of objective measures including Employment Status, Job Level and Years of Service and a subjective Job Performance score which represented each business unit or teams performance ratings against business objectives. The personality inventory was based on the Five Factor Model of Personality and was presented in a Likert scale format. Each of these tests will now be now described in more detail.

3.9.2.1 Digit Sequence Test.

Digit Sequence was also used as a measure of Gf in Study 2. For all full description of this task see Study 1. As a reminder of what this task entails, Figure 2.4 provides an illustration of a Digit Sequence item.

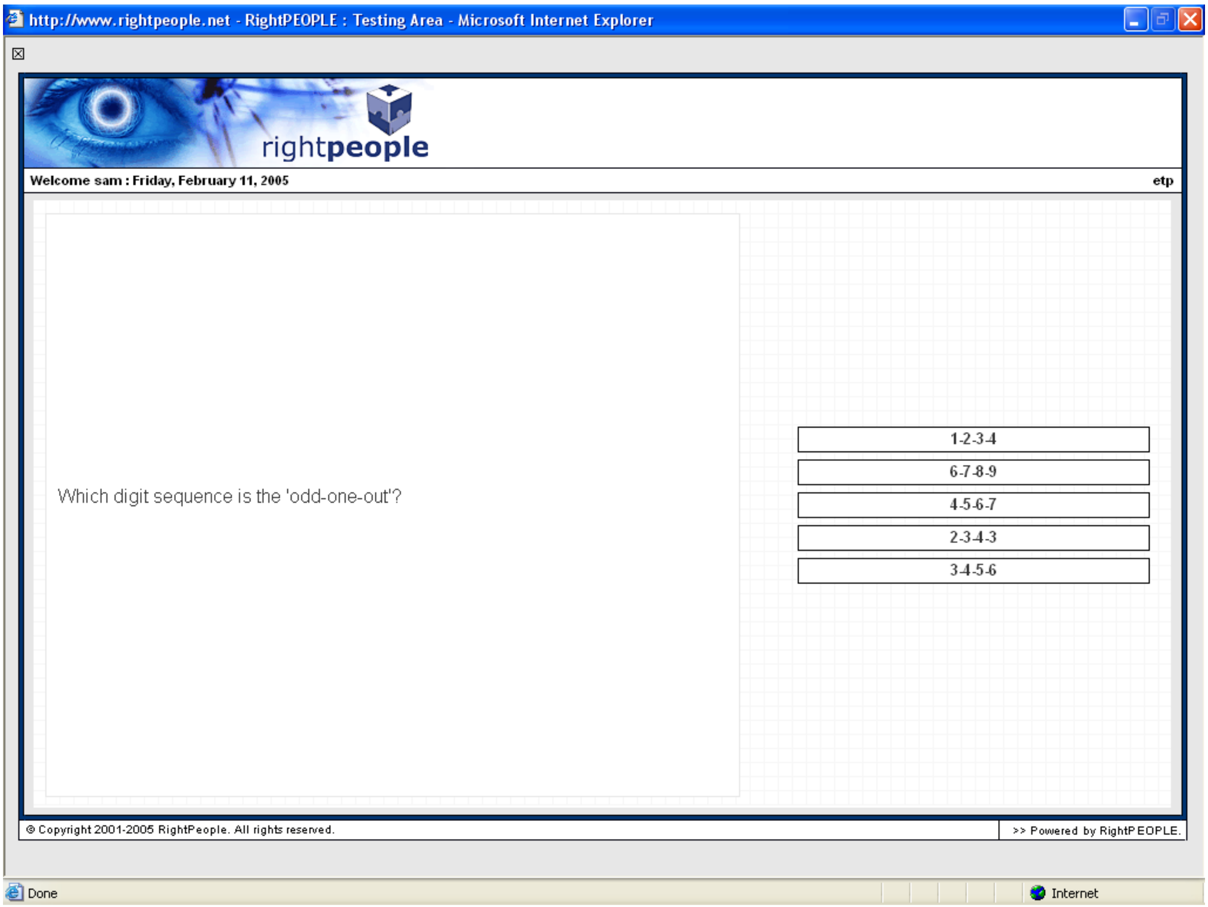


Figure 2.4. Digit sequence task

3.9.2.2 Matrices.

In the Matrices task (source: E-ntelligent Testing Products Pty Ltd), purported to be a measure of Gf, the participant is given an incomplete matrix comprised of a series of circles and crosses arranged in a particular order and must complete it by selecting one of a possible five options by clicking on it with the mouse. Only one of the options will follow the rules of the matrix in terms of the placement of the circles and crosses. There were 12 items in total in the test. There was a time limit of 60 seconds for each question. This is a test of Gf. Responses were scored 0 for incorrect and 1 for correct for each item, leading to a total "Number Correct" score out of 12. Figure 2.5 provides an illustration of a Matrices item.

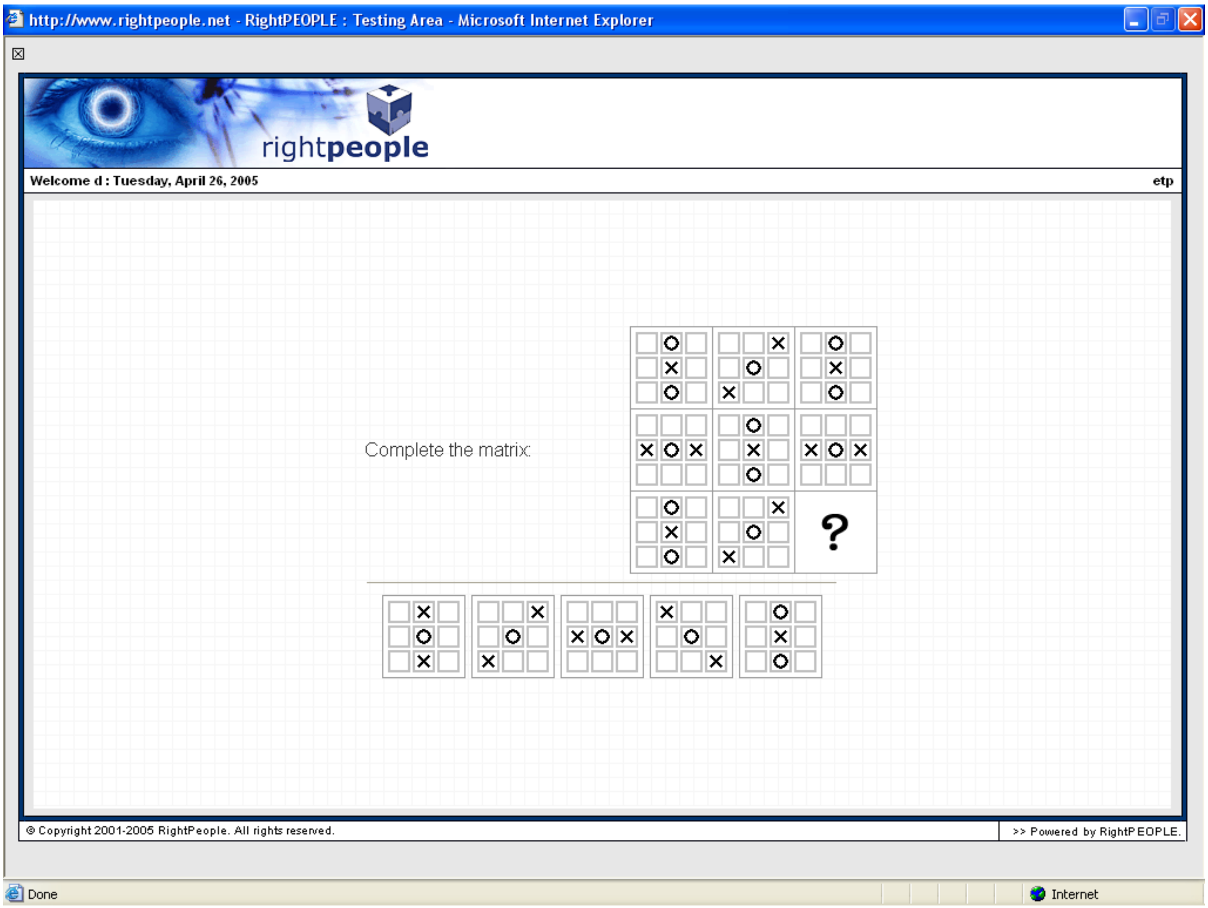


Figure 2.5. Matrices task

3.9.2.3 Reading Comprehension.

In Reading Comprehension (source: E-ntelligent Testing Products Pty Ltd), purported to be a test of crystallized intelligence (Gc), the participant is given a series of vignettes and must decide which of four conclusions provided make sense based on the vignette. They must select the appropriate conclusion by clicking on it with the mouse. There were 20 items in total in the test. Responses were scored 0 for incorrect and 1 for correct for each item, leading to a total “Number Correct” score out of 20. There was a time limit of 120 seconds for each question. Figure 2.6 provides an illustration of a Reading Comprehension item.

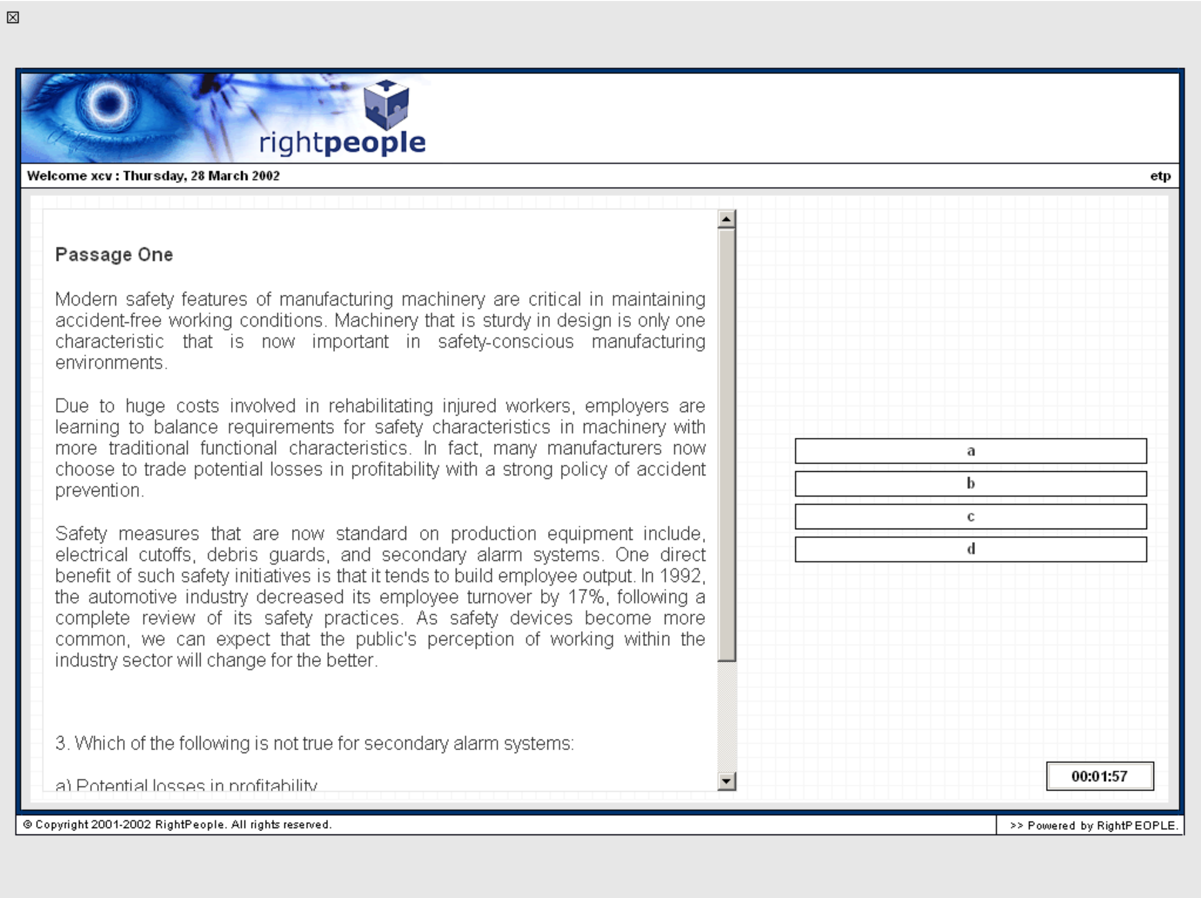


Figure 2.6. Reading Comprehension task

3.9.2.4 Word Reasoning Test.

Word Reasoning was also used as a measure of Gc in Study 2. For all full description of this task see Study 1. As a reminder of what this task entails, Figure 2.7 provides an illustration of a Word Reasoning item.

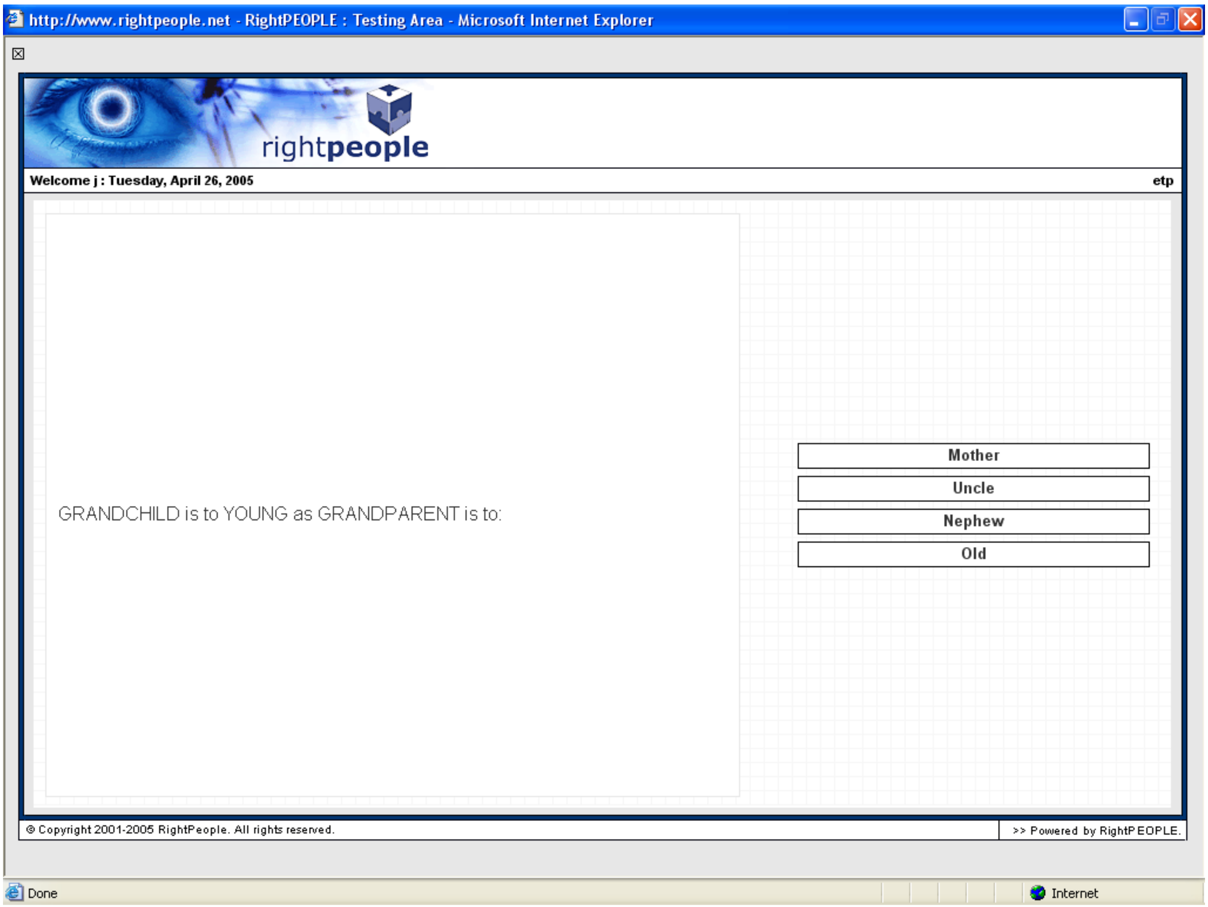


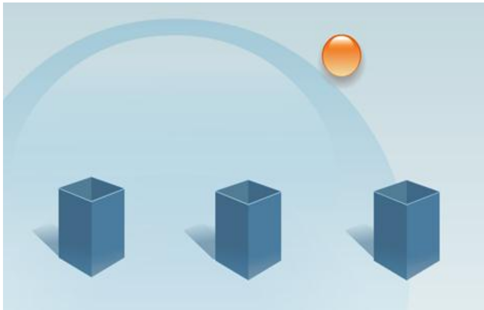
Figure 2.7. Word Reasoning task

3.9.2.5 Digit Span.

In the Digit Span task (source: Fogarty, unpublished), purported to be a measure of Gsm, the participant must remember a series of numbers presented to them in quick succession. The items were presented individually on screen and after the last one was presented the participants must enter the digits in the correct order in the boxes provided. There were 14 items in total. Responses were scored 0 for incorrect and 1 for correct for each item, leading to a total "Number Correct" score out of 14. There was a time limit of 60 seconds for each question. A video demonstration is provided at: http://rightpeople.com.au/researchvideos/ (username: RightPeople, password: Rp1111117). Copy and paste the URL into your Internet browser now.

3.9.2.6 Multi-Tasks Test.

The Multi-Tasks test was administered the same way in Study 2 as in Study 1. For all full description see Study 1. As a reminder of what this task entails, the sub-tasks are illustrated in Figures 2.8 and 2.9.



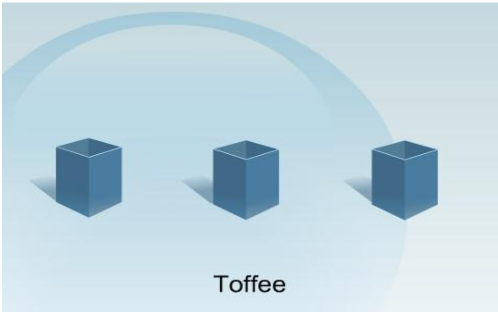


Figure 2.8. Multi-Tasks Placement Keeping component (left) and then Word Recall component (right)

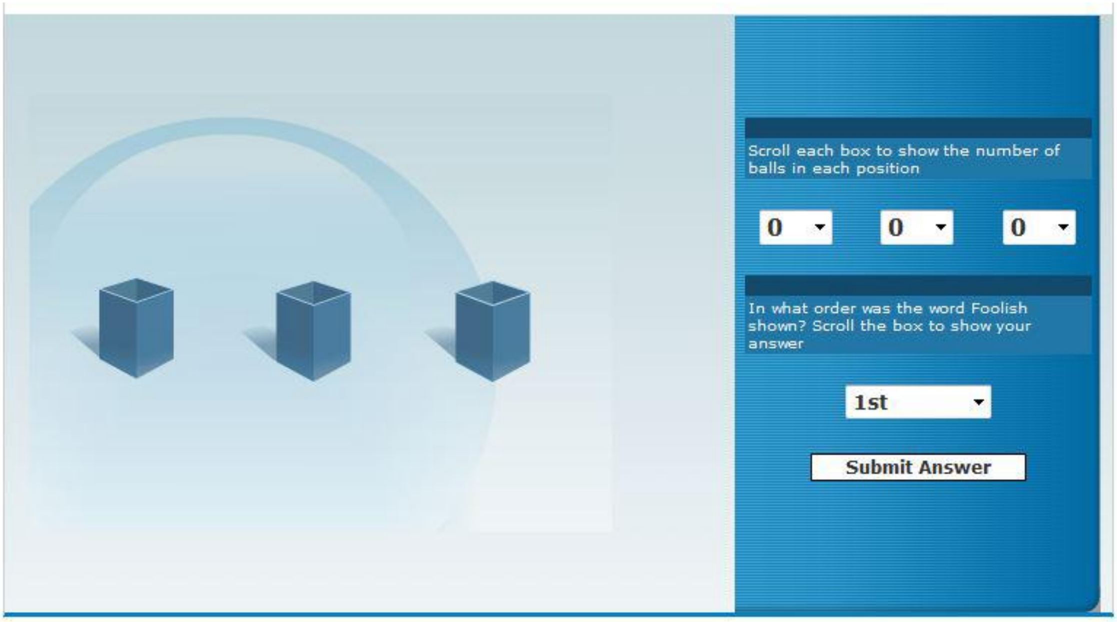


Figure 2.9. Multi-Tasks test answer option whereby the participant must recall the number of balls in each box and then the order of one of the words shown in the display

3.9.2.7 OCEANIC Personality Inventory.

In this measure of personality based on the Five Factor Model, participants were asked to rate their level of agreement on a Likert scale from “never” through to “always” about their preferences and personal style. There were 60 items in total, 12 measuring each of Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Emotional Stability. Items were scored 1 for never through to 6 for always and the overall score for each components was computed by averaging responses to each of the 12 items for each scale. The highest possible total for each scale was therefore 6. There was no time limit for these questions. The OCEANIC has good psychometric properties, and compares favourably with other measures such as the NEO Five Factor Inventory (NEO-FFI, Costa & McCrae, 1992). Internal consistency coefficients range from .77 (O) to .91 (N), compared to .68 (A) to .89 (N) for the NEO-FFI (Costa & McCrae, 1992), significant correlations were found with the NEO-FFI and with external criteria, including university grades, and the underlying model consistent with the FFM (Schulze & Roberts, 2006).

Figure 2.10 provides an example of a question from the personality inventory. Full descriptions of the inventory items are given in Appendix 2E.

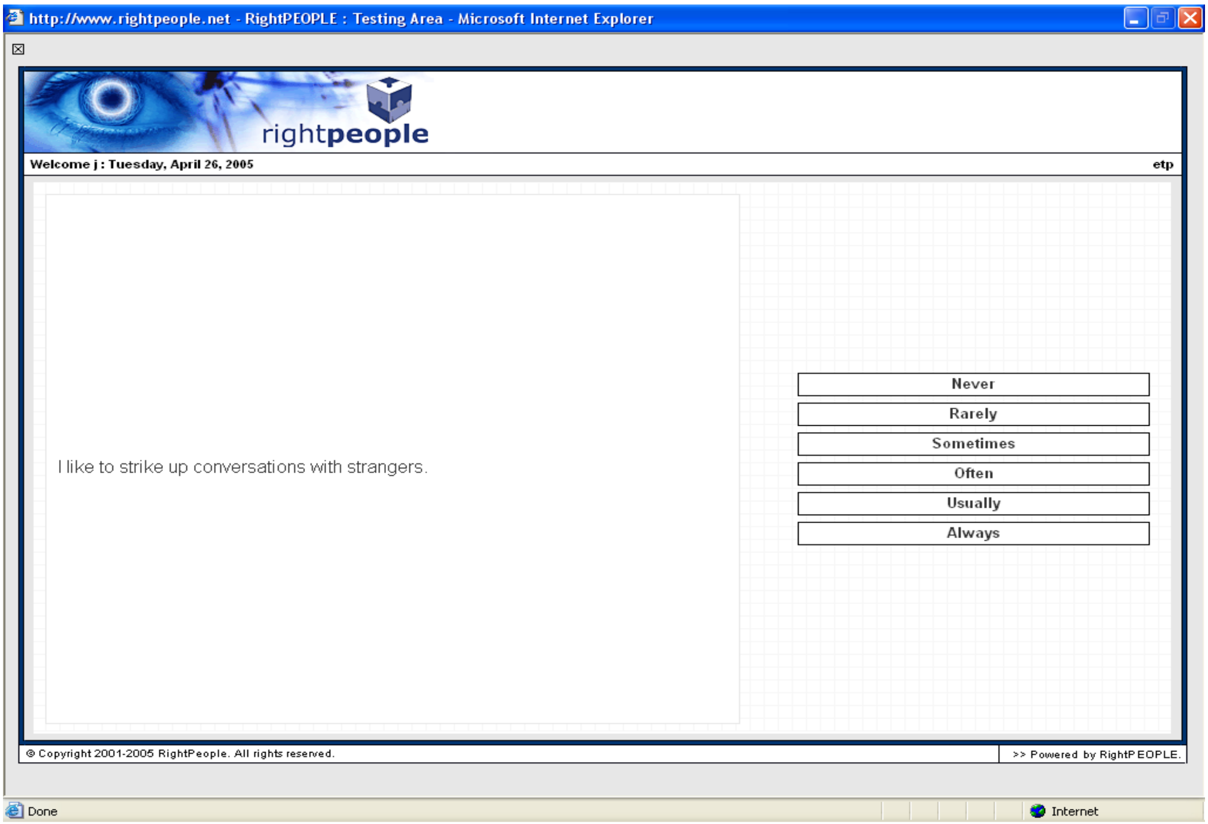


Figure 2.10. Personality Measure

3.9.2.8 Job Performance Measures.

There were four job performance measures. Objective measures included Employment Status (comprising of voluntary transfers of high performing individuals being transferred to more senior positions, and involuntary transfers being those whose performance is unsatisfactory being moved into alternative positions), Job Level (numbered 1-6 with 1 being the most junior and 6 the most senior) and Years of Service (being how long the individual has been employed by the organisation). In addition, some participants had a Job Performance score which represented each business unit or teams performance ratings against business objectives. Objectives were given a rating of 1-4 and each objective had a different weighting. The scores for each objective and their weighting were taking into consideration and the total calculated to give an overall performance score ranging for 0 (lowest) to 3 (highest). These were not consistent between participants as different business units or teams within business units used different objectives to measure performance against.

Ratings were recorded and collated on Excel spreadsheet by the Human Resources Officer. Ratings of employment status were collected for 165 participants, of job level and years of service for 143 participants and for job performance for 50 participants.

3.9.3 Procedure.

Testing was undertaken via the Internet by invitation from the Human Resources Department. The system was supported by Microsoft.net and Adobe Flash on a Microsoft Sequel Server. The system provided the participants with full instructions, including practice questions (2 questions per sub-task) and feedback if the practice questions were completed incorrectly. Each subject was administered the test battery with the same order of sub-tasks. The test order was as follows: Word Reasoning, Reading Comprehension, Multi-Tasks, Matrices, Digit Sequence, Digit Span, and the five personality subscales.

Testing time was approximately 70 minutes. When the tests were completed the system informed the participant that they had completed the assessment. Results were saved to a secure server.

3.10 Results

In the first section, preliminary procedures will be described. In particular, the outlier checks, reliabilities of the measures, and descriptive statistics will be reported. Following this, t-test and analysis of variance results capturing demographic group differences in Multi-Tasks test scores will be described. In the last section, the results of the measurement model and structural equation model tests will be presented.

3.10.1 Preliminary Procedures.

3.10.1.1 Outlier checks.

Total scores were computed for the Multi-Tasks Word Recall, Multi-Tasks Placement Keeping, Matrices, Reading Comprehension, Digit Sequence, Word Reasoning and Digit Span scales, and the five personality subscales, while mean composites were computed for the speed measures. The composites were standardised; cases whose standardised values exceeded the absolute value of 3.29 (per Tabachnick & Fidell, 2007) were considered as outliers. Eighty-five cases had standardised values higher than the absolute value of 3.29 and were deleted from subsequent analyses.

3.10.1.2 Descriptive Statistics and Reliabilities for the Multi-Tasks Subscales.

This section will be structured so that descriptive statistics and reliabilities will be reported separately for the two sub-tasks that comprised each Multi-Tasks measure, Word Recall and Placement Keeping. Note that when items were dropped from the analyses because of their negative effect on alpha levels, they were removed from both the Word Recall and Placement Keeping measures to keep the number of items equal across the measures.

3.10.1.2.1 Word Recall.

The initial Word Recall alpha was .84. Two items were deleted as one had a negative item-total correlation and the other had a low item-total correlation. The resultant alpha was .85 and the mean score was 12.40 (SD = 2.55).

3.10.1.2.2 Placement Keeping.

The items used to measure the Word Recall subscale were used to measure the Placement Keeping subscale. Alpha for this subscale was .88. As there were 14 items retained, the highest Placement Keeping with Balls score was 42. The mean score was 34.12 (SD = 9.08). These statistics are given in Table 2.4.

Table 2.4

Descriptive Statistics and Alpha Coefficients for the Multi-Tasks Subscales (N = 1306)

Subscale	Item N	M	SD	Alpha
Multi-Tasks Word Recall	14	12.40	2.55	.84
Multi-Tasks Placement Keeping	14	34.12	9.08	.88

3.10.1.3 Descriptive Statistics and Reliability Measures for the other Intelligence Scales.

3.10.1.3.1 Matrices.

The initial alpha for this measure was .55. As the item total correlation of one item was .07, it was dropped from the scale. As shown in Table 2.5, alpha after this item was dropped increased to .57. The mean score was 6.85 (SD = 2.05).

3.10.1.3.2 Reading Comprehension.

The initial alpha for the Reading Comprehension scale was .61. Since two items had negative item-total correlations, they were dropped from the scale. Alpha then increased to .66. The mean score was 13.34 (SD = 2.62).

3.10.1.3.3 Digit Sequence.

The initial alpha for the Digit Sequence scale was .36. Three items either had a negative item-total correlation or an item-total correlation less than .10 and were deleted. Alpha for the scale increased to .46 and the mean score was 7.35 (SD = 1.37).

3.10.1.3.4 Word Reasoning.

The initial alpha for the Word Reasoning scale was .64. Two items had item-total correlations below .10 and were dropped from the scale. Thereafter, alpha increased to .66. The mean score was 13.95 (SD = 2.94).

3.10.1.3.5 Digit Span.

The initial alpha for the Digit Span scale was .73. As one item had a negative item-total correlation and four items had item-total correlations ranging from .00 to .03, they were dropped from the scale. Thereafter, alpha increased to .77. The mean score was 7.56 (SD = 1.94).

These statistics are given in Table 2.5.

Table 2.5

Descriptive Statistics and Alpha Coefficients for the Intelligence Scales (N = 1306)

Subscale	Item N	M	SD	Alpha
Matrices	11	6.85	2.08	.57
Reading Comprehension	18	13.34	2.62	.61
Digit Sequence	9	7.35	1.37	.46
Word Reasoning	20	13.95	2.94	.65
Digit Span	9	7.56	1.94	.77

3.10.1.4 Descriptive Statistics for the Five Personality Subscales.

Descriptive statistics are given in Table 2.6.

Table 2.6

Descriptive Statistics and Alpha Coefficients for the Personality Subscales (N = 1306)

Subscale	Item N	Range	M	SD	Alpha
Openness	12	6	3.57	.69	.83
Conscientiousness	12	6	5.30	.52	.87
Extraversion	12	6	4.60	.60	.83
Agreeableness	12	6	5.48	.44	.86
Neuroticism	12	6	1.79	.45	.86

3.10.1.5 Descriptive Statistics for the Job Performance Measures.

The frequencies and percentages for the job performance variables measured on a nominal or ordinal scale are displayed in Table 2.7. Of the 165 respondents whose employment status was assessed, the majority were voluntary transfers (89.1%). Of the 143 respondents whose job level information was available, the majority were at Level 1 (86.7%);

a minority were at Level 2 (11.2%) or at Level 3 (2.1%); none of the respondents were at Levels 4 to 6.

Table 2.7

Frequencies and Percentages for Job Performance Measures

Variable	Frequency	Percentage
Employment status		
Voluntary transfer	147	89.1
Involuntary transfer	18	10.9
Job level		
One	124	89.7
Two	16	11.2
Three	3	2.1

As shown in Table 2.8, respondents had between one month and 20 years of service with the organisation. The mean number of years was .91 (SD = 2.10). Job performance scores ranged from zero to three; the mean job performance score was 1.64 (SD = .97).

Table 2.8

Descriptive Statistics for Job Performance Measures

Measure	N	M	SD
Years of service	143	.91	2.09
Job performance score	50	1.64	.97

3.10.1.6 Correlations between variables.

As most of the variables were not normally distributed (cf. the histograms in Appendix 2A), Kendall’s tau correlations between the study variables are reported in Table 2.9. The findings reveal that the Multi-Tasks Word Recall subscale was positively associated with the Matrices scale (τ = .11, p = .001), Digit Sequence (τ = .08, p = .001), Digit Span (τ = .38, p = .001), Word Reasoning (τ = .06, p = .003), and less so with Reading Comprehension (τ = .04, p = .048).

The Multi-Tasks Placement Keeping subscale was positively correlated with the Matrices scale (τ = .17, p = .001), Reading Comprehension (τ = .14, p = .001), Digit Sequence (τ = .19, p = .001), Word Reasoning (τ = .20, p = .001), and Digit Span (τ = .34, p = .001).

Table 2.9

Kendall Tau Correlations between Study Variables (N = 1306)

Variable	1	2	3	4	5	6
1 Multi-Tasks word recall						
2 Multi-Tasks placement keeping	.48***					
3 Matrices score	.11***	.17***				
4 Reading comprehension score	.04*	.14***	.22***			
5 Digit sequence score	.08***	.19***	.19***	.20***		
6 Word reasoning score	.06**	.20***	.22***	.32***	.16***	
7 Digit span score	.38***	.34***	.17***	.01	.03	.10**

*p < .05. **p < .01. ***p < .001.

3.10.2 Multi-Tasks Scores across Demographic Groups.

Independent t-test and one-way analysis of variance (ANOVA) procedures were conducted to determine whether scores varied across gender, levels of education, and age. A two-tailed significance level was specified for the independent t-test procedure. A significance level of .05 was specified for the overall model in the ANOVA procedure; when the overall model was statistically significant, post-hoc Tukey procedures were conducted.

3.10.2.1 Gender.

The means and standard deviations for the Multi-task tests are presented in Table 2.10. Results indicate that the Multi-Tasks Word Recall scores of males did not significantly differ from that of females, t (1302) = .26, p = .793; r^2 = .000. Similarly, the Multi-Tasks Placement Keeping scores of males did not differ significantly from that of females, t (802) = 1.41, p = .158; r^2 = .002.

Table 2.10

Means and Standard Deviations for Multi-Tasks Scores of Males and Females

Gender	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Males	430	12.32	2.50	33.56	9.74
Females	874	12.44	2.58	34.40	8.88

3.10.2.2 Level of Education.

The means and standard deviations for the Multi-task tests are displayed in Table 2.11. The findings revealed that Multi-Tasks Word Recall scores did not vary significantly across levels of education, F (5, 1254) = .85, p = .517; η^2 = .004, but Multi-Tasks Placement Keeping scores did differ significantly across levels of education, F (5, 1254) = 2.59, p = .024; η^2 = .011. Post-hoc Tukey tests, however, did not yield any significant group comparisons.

Table 2.11

Means and Standard Deviations for Multi-Tasks Scores across Levels of Education

Level of Education	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
Secondary certificate	252	12.51	2.35	34.37	8.58
Higher school certificate	430	12.50	2.38	33.72	9.88
TAFE qualified	198	12.10	2.90	33.23	9.98
Took college courses	103	12.29	2.57	34.26	8.67
Undergraduate degree	215	12.53	2.48	35.53	7.96
Postgraduate degree	62	12.26	2.53	31.50	11.35

3.10.2.3 Age.

The means and standard deviations for the Multi-Tasks tests are shown in Table 2.12. The findings indicate that Multi-Tasks Word Recall scores did not vary across age groups, $F (7, 1296) = 1.03$, $p = .408$; $\eta^2 = .000$. Similarly, Multi-Tasks Placement Keeping scores did not vary significantly across age groups, $F(7, 1296) = .19$, $p = .988$; $\eta^2 = .000$.

Table 2.12

Means and Standard Deviations for Multi-Tasks Scores across Age Groups

Age Group	N	Word Recall		Placement Keeping	
		Mean	SD	Mean	SD
18 to 24	559	12.23	2.81	33.99	9.39
25 to 29	337	12.50	2.30	33.96	9.19
30 to 34	140	12.64	2.07	34.35	8.99
35 to 39	79	12.34	2.56	34.24	8.96
40 to 44	67	12.34	3.07	34.42	7.85
45 to 49	57	12.74	2.23	35.26	7.33
50 to 54	36	12.42	2.52	33.97	9.34
55 and above	29	13.03	1.30	34.24	8.74

3.10.3 Results of Measurement and Structural Model Tests.

3.10.3.1 Parcels for Measurement and Structural Model Tests.

Parcels (i.e., small item groups) were created for two reasons. Firstly, the tests were scored using a nominal scale (i.e., incorrect vs. correct) and because a requirement of structural equation modeling is that the data be measured on either an interval or ratio scale, items had to be summed into parcels. Secondly, as Little, et al (2002) identified, models with single-item indicators are less parsimonious and often increase sampling error. Therefore, parcels were created.

3.10.3.1.1 Multi-Tasks Parcels.

The Multi-Tasks items were arranged in order of difficulty. Parcels were thus created based on number order. Four parcels consisting of three items were created; the first parcel had the least difficult items while the fourth parcel had the most difficult items (refer to Appendix 2B for the parcel compositions).

3.10.3.1.2 Parcels for the other Intelligence Test Scales.

As items for these tests were not arranged in order of difficulty, the item-to-construct balance method, recommended by Little, et al. (2002) was used to create parcels. Specifically, the corrected item-total correlations (from the reliability analyses) were used to generate the parcels. These correlations were sorted from highest to lowest. Items with the highest item-total correlation anchored each of the parcels. The items with the next highest item-total correlations were added to the anchors in the reverse order. The item with the highest correlation among the anchor items was matched with the lowest loading item from the second selection. This basic procedure where items with lower correlations were matched with items with higher correlations was repeated until all items were categorised into parcels. Refer to Appendix 2B for the parcel compositions.

3.10.3.1.2.1 Normality of Parcels.

Multivariate normality is an important assumption of structural equation modelling, thus normality of the parcels was assessed. According to Kline (2005), multivariate normality can be assumed when the univariate distributions are normal and the distribution of any pair of variables is bivariate normal. Kline further proposed that because it is impractical to examine all joint distributions, examining univariate distributions will usually allow one to detect instances of non-normality. Thus, the skewness and kurtosis of the parcels were assessed.

Most of the variables were highly skewed; their skew indices were above the acceptable criterion of three (Kline, 2005). Accordingly, these variables were transformed to correct for skewness (Howell, 1992). The variables that were negatively skewed were normalised using a power transformation; the variables that were positively skewed were normalised using a square root transformation. These transformed variables were used in subsequent analyses.

3.10.3.2 Model Evaluation.

Model fit was assessed by interpreting several fit indices including the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), the Standardised root mean residual (SRMR), and the likelihood ratio χ^2 test. A model is deemed as fitting the data well when the CFI value is above .95 (Hu & Bentler, 1999). Browne and Cudeck (1993) suggest that a model with an RMSEA value less than .05 has good fit, one with a value than .08 has reasonable fit, and a model with an RMSEA less than .10 has poor fit. Kline (2005) revealed that a model with an SRMR value of less than .10 has good fit. A small χ^2 value relative to the degrees of freedom (i.e., values lower than 3) indicates good model fit (Hu & Bentler, 1999).

Nested models were compared via the change in chi-square. In addition, two information indices were reported: the Akaike Information Criterion (AIC) and the Expected Cross-Validation Index (ECVI). Information indices are relative; models indices are compared to each other and the lower the value, the better the model fit (Byrne, 2001).

In addition to evaluating the model as a whole, the significance of the individual parameters was also assessed (Byrne, 2001). Parameters were evaluated at the .05 level. The direction of the standardised path coefficients was checked to see if it was consistent with expectations.

3.10.3.3 Confirmatory Factor Analysis of Multi-Tasks Tests.

The proposed two-factor model using the transformed Multi-task parcels had mediocre fit. To improve model fit, the first Placement Keeping and Word Recall parcels were dropped, as their standardized factor loadings were low in comparison to the loadings of the other parcels (Figure 2.11). Model fit for this revised two-factor model was adequate, with some of the fit statistics meeting the criteria outlines earlier (e.g., CFI and SRMR). This revised two-factor model was compared to a single-factor model. The fit indices for the models are summarised in Table 2.13. The findings revealed that the two-factor model was the best fit. Further, the change in chi-square between the single-factor and two-factor model was statistically significant ($\Delta\chi^2$ (2) = 282.88, $p < .001$). Lastly, values of the information indices for the two-factor model were lower than the values for the single- factor model. Maximum likelihood estimates for the variables, and correlations among the factors are shown in Appendix 2C. This two-factor model was carried through to the next stage of the analysis.

Table 2.13

Fit Indices for the Multi-Tasks Measurement Models

Index	Two Factor	Revised Two Factor	Single Factor
Chi-square	357.56	86.59	369.47
Degrees of freedom	19.00	8.00	9.00
Sig.	.00	.00	.00
Chi-square/df	18.82	10.82	41.05
Comparative fit index (CFI)	.92	.97	.88
Root mean squared error (RMSEA)	.12	.09	.18
Lower bound 90 % CI	.11	.07	.16
Upper bound 90 % CI	.13	.10	.19
Standardised root mean residual (SRMR)	.05	.03	.07
Akaike information criterion (AIC)	391.56	112.59	393.47
Expected cross-validation index (ECVI)	.30	.09	.30

Note. At $p < .001$, critical χ^2_{crit} (2) = 13.82.

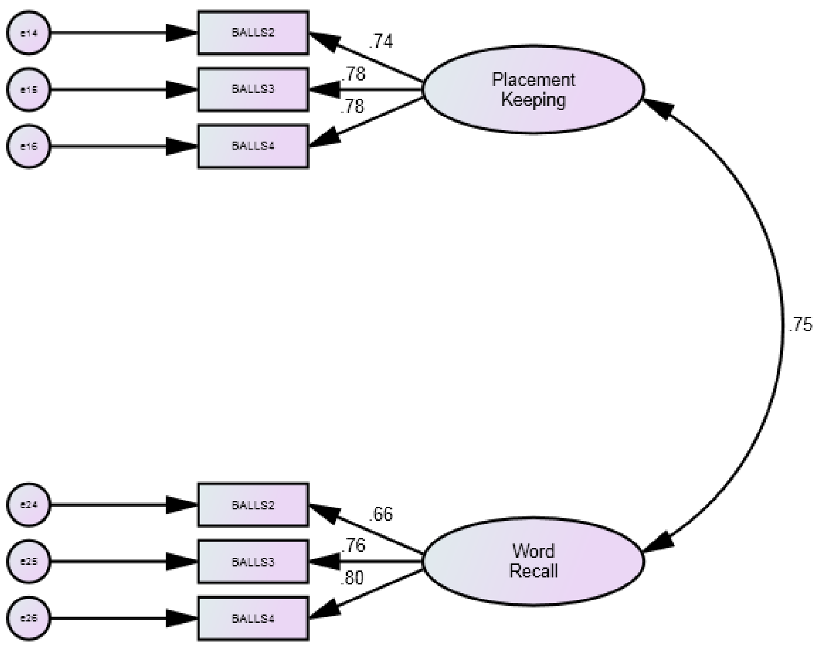


Figure 2.11. Revised two-factor model for multi-tasks test (standardised coefficients)

3.10.3.4 Confirmatory Factor Analysis of the Full Measurement Model.

Prior to testing the proposed full measurement model, a first-order confirmatory factor analysis was conducted. Indicator variables whose standardised coefficients were less than .50 (per Hair, et al., 2010) were deleted from the model. The following indicator variables were thus deleted: Digit Sequence 3 (β = .41) and Read 2 (β = .48).

The proposed measurement models are depicted in Figure 2.12 and Figure 2.13. The fit indices for these models are summarised in Table 2.14. The two-factor model (Proposed Model 1) did not fit the data well. In addition, the error variance of the Word Reasoning construct was negative. Similarly, the three-factor model (Proposed Model 2) did not fit the data well. Accordingly, the two-factor proposed model was revised. Firstly, the standardised coefficients from the second-order factors to the first-order factors were examined. As the Digit Sequence and Matrices loadings were also low, they were made to load onto the Fluid Intelligence construct. Secondly, the modification indices were examined. Indicator variables with high modification indices (MI) were deleted as this was an indication that the variables did not have discriminant validity (Byrne, 2001). Accordingly, the following indicator variables were deleted: Matrix 3 (MI for loading onto MT Balls Word Recall = 21.95), Read 3 (MI for loading onto MT Balls Word Recall 4 = 26.83), and Reason 3 (MI for the correlation of its error variance with the error variance of MT Word Recall = 40.64). Lastly, because the items measuring the Multi-Tasks Word Recall and Placement Keeping parcels were the same, correlations between the error terms of the indicator variables were added.

The revised model of intelligence is depicted in Figure 2.14 and the fit indices are summarised in Table 2.14. Path coefficients are given in Appendix 2D. This revised model fitted the data well. It also had significantly better fit than the proposed two- and three-factor models: the changes in chi-square were statistically significant ($\Delta\chi^2(54) = 600.47$, $p < .001$; $\Delta\chi^2(52) = 346.13$, $p < .001$) and the values of the information indices of the revised model were lower than those of the proposed models.

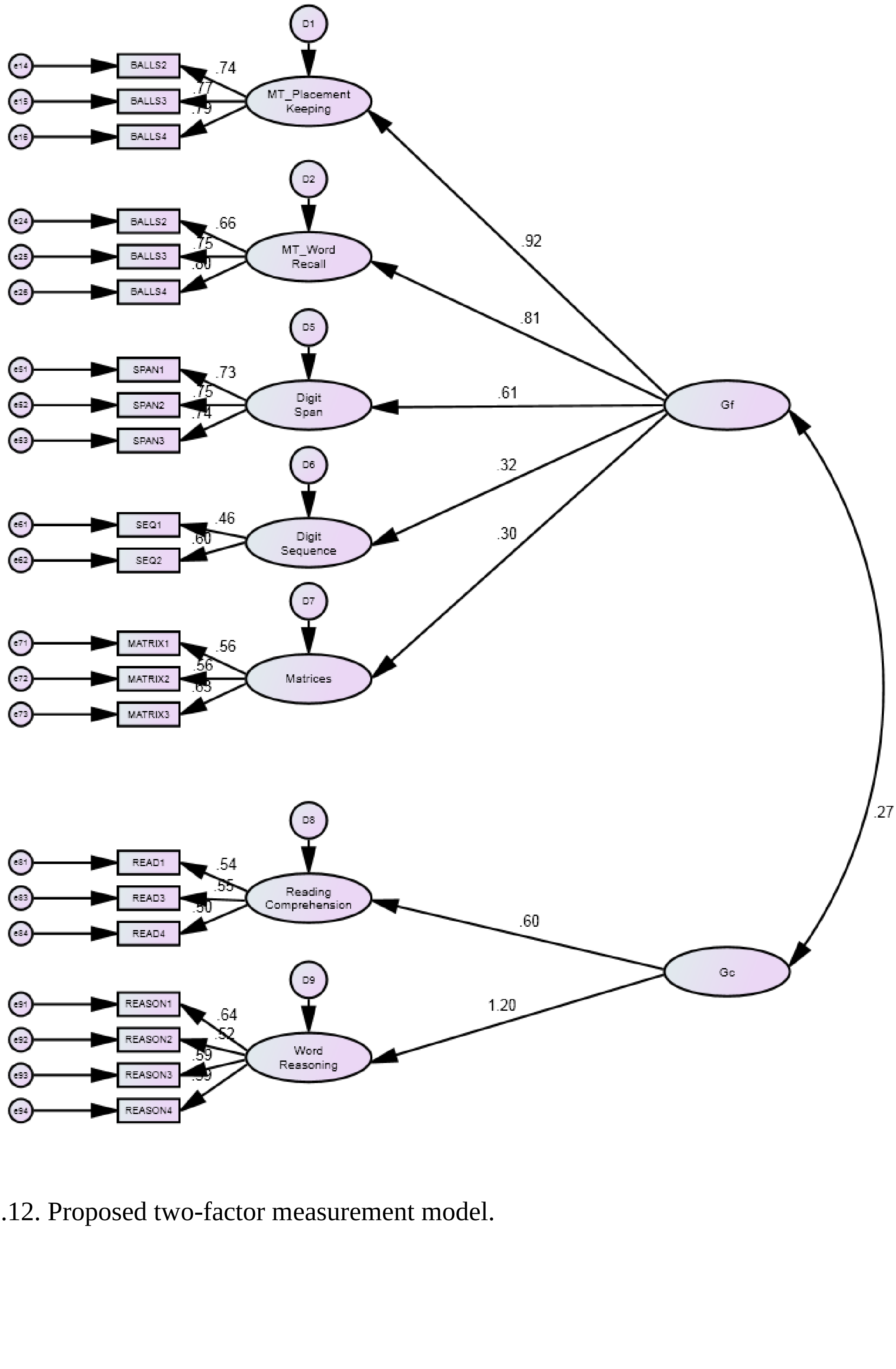


Figure 2.12. Proposed two-factor measurement model.

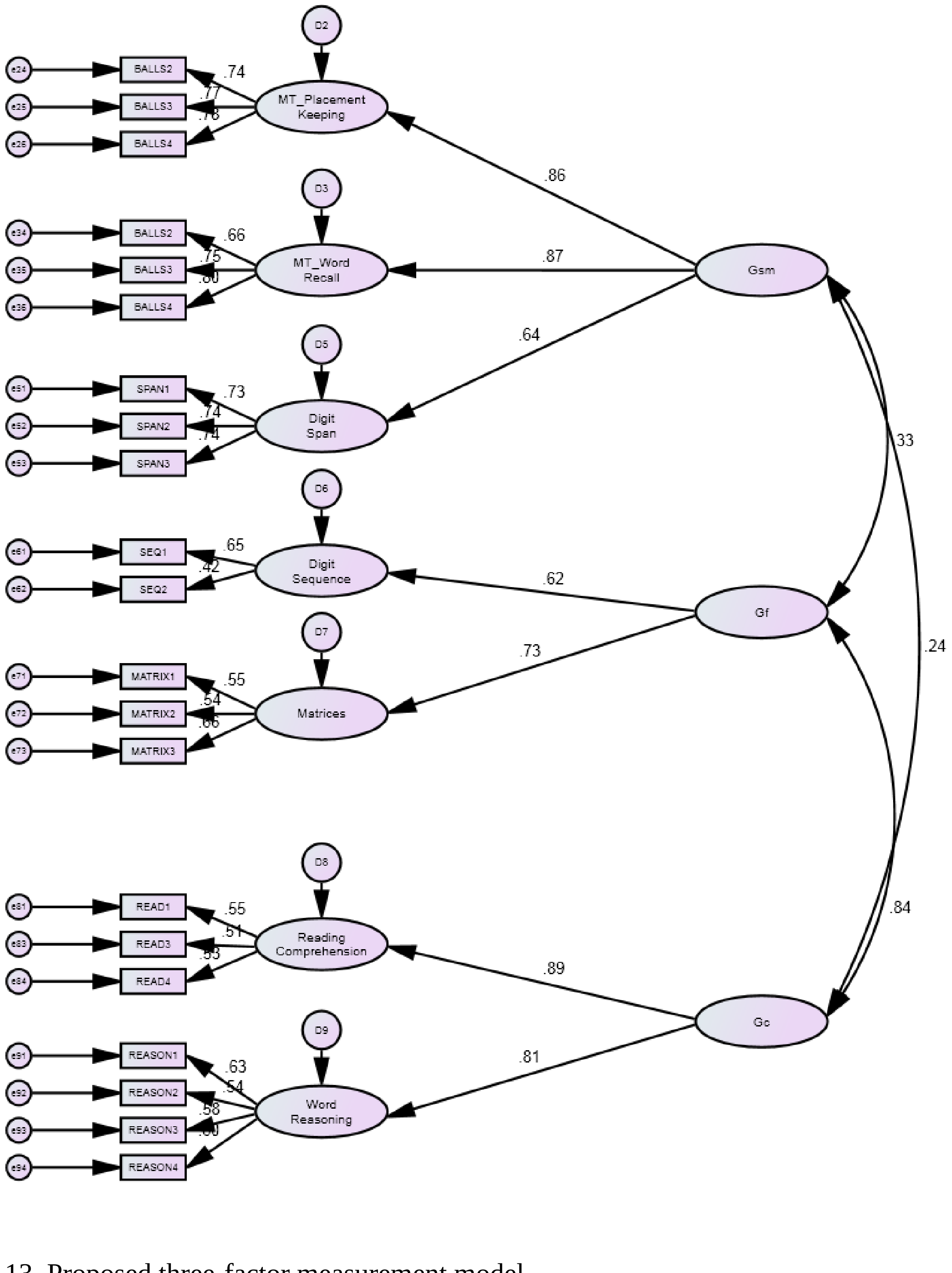


Figure 2.13. Proposed three-factor measurement model.

Table 2.14

Fit Indices for the Intelligence Model

	Proposed Two Factor	Proposed Three Factor	Revised Two Factor
Chi-square	1219.36	965.02	618.89
Degrees of freedom	181.00	179.00	127.00
Sig.	.00	.00	.00
Chi-square/df	6.74	5.39	4.87
Comparative fit index (CFI)	.86	.89	.92
Root mean squared error (RMSEA)	.07	.06	.05
Lower bound 90 % CI	.06	.05	.05
Upper bound 90 % CI	.07	.06	.06
Standardised root mean residual (SRMR)	.07	.05	.04
Akaike information criterion (AIC)	1319.36	1069.02	706.89
Expected cross-validation index (ECVI	1.01	.82	.54

Note. At $p < .001$, critical χ^2_{crit} (54) = 91.87 and critical χ^2_{crit} (52) = 89.27

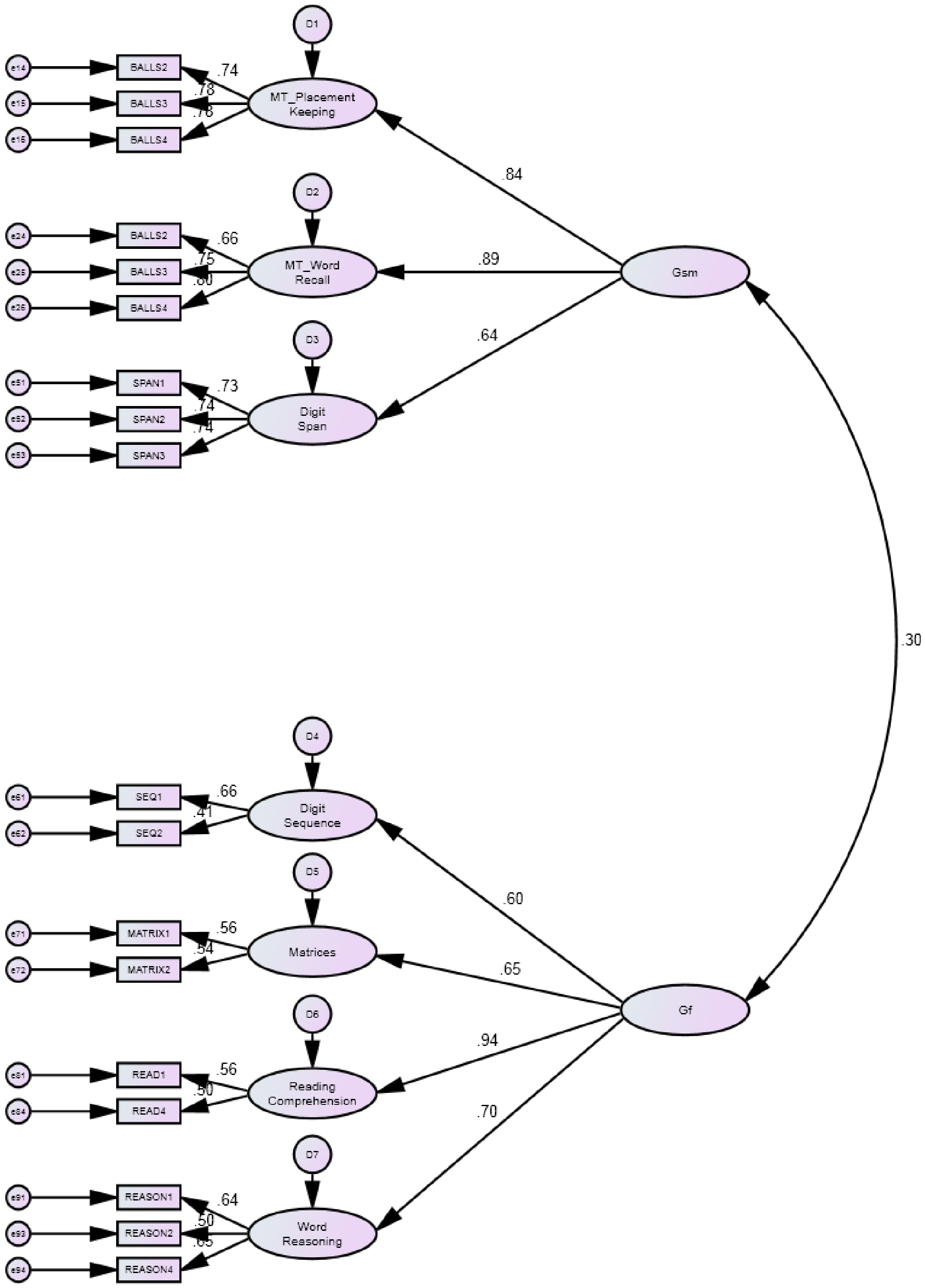


Figure 2.14. Revised measurement model.

3.10.3.5 Predictive Validity of the Intelligence Measures.

To determine the predictive validity of the Multi-Tasks test vis-à-vis the other intelligence tests, Pearson correlations and several regression procedures were conducted. Linear regression procedures were conducted with the job performance and the transformed years of service variables because these variables were measured either on an interval or ratio scale and were distributed normally. Logistic regression procedures were conducted with the job level (level 1 vs. levels 2 and 3) and employment status (voluntary vs. involuntary transfer) variables.

3.10.3.5.1 Job Performance.

The Pearson correlations between the Multi-Tasks subscales, the other intelligence measures, and job performance scores are presented in Table 2.15. The findings reveal that none of the intelligence measures were significantly correlated with job performance scores.

Table 2.15

Pearson Correlations between Intelligence Measures and Job Performance (N = 50)

Test	r with JP	Sig.	r^2
Multi-Tasks word recall	-.05	.708	.003
Multi-Tasks placement keeping	-.01	.947	.000
Matrices	-.11	.460	.012
Reading comprehension	.05	.748	.003
Digit sequence	.11	.443	.012
Word reasoning	-.13	.358	.017
Digit span	.04	.795	.002

Note. JP = Job Performance.

The linear regression results for the job performance model are summarised in Table 2.16. The findings revealed that none of the tests significantly predicted job performance.

Table 2.16

Linear Regression Results for the Job Performance Model (N = 50)

Test	β	F	Sig.	r^2	Tolerance
Multi-Tasks word recall	-.08	.17	. 680	.006	.58
Multi-Tasks placement keeping	.10	.24	.626	.010	.51
Matrices	-.10	.23	.633	.010	.50
Reading comprehension	.40	2.38	.131	.160	.32
Digit sequence	.08	.21	.649	.006	.74
Word reasoning	-.46	2.66	.110	.211	.26
Digit span	.15	.63	.431	.023	.60

Note. Overall model F (7, 42) = .73, p = .649. Overall model R^2 = .108.

3.10.3.5.2 Years of service.

The Pearson correlations between the Multi-Tasks subscales, the other intelligence measures, and years of service (transformed) are presented in Table 2.17. The findings reveal that the Multi-Tasks Word Recall measure was positively correlated with years of service (r = .23, p = .005). The Multi-Tasks Placement Keeping test was also positively correlated with years of service (r = .22, p = .009). In addition, the Digit Sequence (r = .18, p = .031) and Digit Span (r = .19, p = .036) tests were positively correlated with years of service. Therefore, the higher the Multi-Tasks (Word Recall and Placement Keeping), Digit Sequence, and Digit Spans scores, the longer the respondents spent working in the organization (note that as age was not correlated with years of service, r = -.04, p = .650, its effect was not partialed out).

Table 2.17

Pearson Correlations between Intelligence Measures and Years of Service (N = 143)

Test	r with YOS	Sig.	r^2
Multi-Tasks word recall	.23	.005	.053
Multi-Tasks placement keeping	.22	.009	.048
Matrices	.08	.354	.006
Reading comprehension	.12	.171	.014
Digit sequence	.18	.031	.032
Word reasoning	.13	.131	.017
Digit span	.19	.026	.036

Note. YOS = Years of Service.

The linear regression results for the years of service model are summarised in Table 2.18. The findings reveal that, after controlling for performance on the other intelligence tests, none of the tests significantly predicted years of service.

Table 2.18

Linear Regression Results for the Years of Service Model (N = 143)

Test	β	F	Sig.	r^2	Tolerance
Multi-Tasks word recall	.15	1.57	.213	.023	.48
Multi-Tasks placement keeping	.07	.51	.476	.005	.62
Matrices	-.03	.13	.722	.000	.76
Reading comprehension	.03	.05	.818	.000	.57
Digit sequence	.14	2.58	.111	.019	.85
Word reasoning	.06	.32	.572	.004	.55
Digit span	.06	.32	.574	.004	.54

Note. Overall model F (7, 135) = 2.14, p = .043. Overall model R^2 = .100.

3.10.3.5.3 Employment Status.

The logistic regression results for the employment status model are displayed in Table 2.19. The findings reveal that the Multi-Tasks Word Recall subscale significantly predicted the likelihood of voluntary transfer (B = .25, p = .012). As Word Recall scores increased, the likelihood of voluntary transfer increased by 1.25.

Table 2.19

Logistic Regression Results for the Employment Status Model (N = 165)

Test	B	SE	Wald	Sig.	OR
Multi-Tasks word recall	.25	.10	6.25	.012	1.28
Multi-Tasks placement keeping	-.05	.04	1.86	.173	.95
Matrices	-.08	.15	.30	.586	.92
Reading comprehension	.14	.12	1.48	.224	1.16
Digit sequence	.04	.20	.04	.842	1.04
Word reasoning	.01	.11	.01	.935	1.10
Digit span	-.10	.14	.58	.445	.90

Note. OR = Odds Ratio. Overall model χ^2 (7) = 9.95, p = .191.

3.10.3.5.4 Job Level.

The logistic regression results for the job level model are displayed in Table 2.20. The participants were at levels 1, 2 or 3. The findings reveal that the Multi-Tasks Word Recall subscale significantly predicted the odds of being categorised in a higher job level (B = -.72, p = .005). As Word Recall scores increased, the odds of being categorised in a job level above one decreased by .48. The Multi-Tasks Placement Keeping subscale also significantly predicted the odds of being categorised in a higher job level (B = .50, p = .002). As Placement Keeping scores increased, the odds of being categorised in a job level above one increased by 1.65. Lastly, the Matrices scale significantly predicted the odds of being categorised in a higher job level (B = .52, p = .022). As Matrices scores increased, the odds of being categorised in a job level above one increased by 1.69. Note that as age was not correlated with job level, r = -.043, p = .677, its effect was not partialed out.

Table 2.20

Logistic Regression Results for the Job Level Model (N = 143)

Test	B	SE	Wald	Sig.	OR
Multi-Tasks word recall	-.75	.27	7.80	.005	.48
Multi-Tasks placement keeping	.50	.16	9.32	.002	1.65
Matrices	.52	.23	5.28	.022	1.69
Reading comprehension	-.27	.18	2.20	.138	.76
Digit sequence	.77	.37	4.41	.036	2.16
Word reasoning	.19	.16	1.36	.243	1.21
Digit span	.53	.28	3.63	.057	1.70

Note. OR = Odds Ratio. Overall model χ^2 (7) = 57.16, p = .001.

3.10.3.6 Predictive Validity of the Multi-Tasks and Personality Measures.

The Pearson correlations between the five personality subscales are summarised in Table 2.21. The findings reveal that the Openness, Conscientiousness, Extraversion, and Agreeableness subscales were all positively associated with each other. The four subscales were all negatively associated with the Neuroticism subscale.

Table 2.21

Pearson Correlations amongst the Personality Subscales (N = 1306)

Variable	1	2	3	4
1 Openness				
2 Conscientiousness	$.26^{***}$			
3 Extraversion	$.25^{***}$	$.27^{***}$		
4 Agreeableness	$.22^{***}$	$.53^{***}$	$.28^{***}$	
5 Neuroticism	$-.07^{*}$	$-.39^{***}$	$-.36^{***}$	$-.39^{***}$

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

The Pearson correlations between the Multi-tasks measures and the five personality subscales are presented in Table 2.22. The findings indicate that the Multi-tasks measures were not significantly correlated with any of the personality subscales.

Table 2.22

Pearson Correlations between the Multi-Tasks test and the Personality Subscales (N = 1306)

Variable	MT Word Recall	MT Placement Keeping
Openness	.00	.00
Conscientiousness	-.02	-.02
Extraversion	-.01	.02
Agreeableness	-.02	.00
Neuroticism	.00	.01

To determine the predictive validity of the Multi-Tasks test vis-à-vis the five personality traits, several regression procedures were conducted. Linear regression procedures were conducted with the job performance and the transformed years of service variables because these variables were measured either on an interval or ratio scale and were distributed normally. Logistic regression procedures were conducted with the job level (level 1 vs. levels 2 and 3) and employment status (voluntary vs. involuntary transfer) variables.

3.10.3.6.1 Job Performance.

The linear regression results for the job performance model are summarised in Table 2.23. The findings reveal that none of the tests significantly predicted job performance.

Table 2.23

Linear Regression Results for Multi-Tasks, Personality Traits, and Job Performance (N = 50)

Test	β	F	Sig.	Tolerance
Openness	-.17	1.23	.275	.89
Conscientiousness	-.08	.16	.690	.57
Extraversion	.20	1.12	.295	.63
Agreeableness	.11	.38	.543	.70
Neuroticism	-.08	.14	.715	.50
Multi-Tasks word recall	-.06	.11	.738	.75
Multi-Tasks placement keeping	.03	.02	.884	.73

Note. Overall model F (7, 42) = .63, p = .727. Overall model R^2 = .095.

3.10.3.6.2 Years of Service.

The linear regression results for the years of service model are summarised in Table 2.24. The findings reveal that none of the tests significantly predicted years of service. Note, however, that the Multi-Tasks Word Recall subscale marginally and positively predicted years of service (β = .17, p = .077).

Table 2.24

Linear Regression Results for Multi-Tasks, Personality Traits, and Years of Service (N = 143)

Test	β	F	Sig.	Tolerance
Openness	.05	.34	.561	.91
Conscientiousness	.04	.18	.675	.64
Extraversion	-.00	.00	.985	.71
Agreeableness	-.06	.31	.580	.68
Neuroticism	-.10	.84	.362	.61
Multi-Tasks word recall	.17	3.17	.077	.78
Multi-Tasks placement keeping	.15	2.43	.122	.79

Note. Overall model F (7, 135) = 1.77, p = .099. Overall model R^2 = .084.

3.10.3.6.3 Employment Status.

The logistic regression results for the employment status model are displayed in Table 2.25. The findings reveal that the Multi-Tasks Word Recall subscale significantly predicted the likelihood of voluntary transfer (B = .24, p = .012). As Word Recall scores increased, the likelihood of voluntary transfer increased by 1.28.

Table 2.25

Logistic Regression Results for Multi-Tasks, Personality Traits, and Employment Status (N = 165)

Test	B	SE	Wald	Sig.	OR
Openness	-.50	.39	1.58	.208	.61
Conscientiousness	1.08	.71	2.35	.127	2.94
Extraversion	-.41	.58	.49	.483	.67
Agreeableness	.29	.73	.16	.692	1.34
Neuroticism	.64	.74	.78	.378	1.93
Multi-Tasks word recall	.21	.08	6.77	.009	1.23
Multi-Tasks placement keeping	-.05	.03	2.31	.128	.95

Note. OR = Odds Ratio. Overall model χ^2 (7) = 12.01, p = .100.

3.10.3.6.4 Job Level.

The logistic regression results for the job level model are displayed in Table 2.26. The findings reveal that the Multi-Tasks Placement Keeping subscale significantly predicted the odds of being categorised in a higher job level (B = .40, p = .000). As Placement Keeping scores increased, the odds of being categorised in a job level above one increased by 1.49.

Table 2.26

Logistic Regression Results for Multi-Tasks, Personality Traits, and Job Level (N = 143)

Test	β	SE	Wald	Sig.	OR
Openness	-.21	.46	.21	.643	.81
Conscientiousness	1.34	1.00	1.78	.182	3.82
Extraversion	.18	.64	.07	.785	1.19
Agreeableness	-.78	.89	.78	.378	.46
Neuroticism	-.86	.88	.94	.332	.42
Multi-Tasks word recall	-.28	.20	1.99	.159	.75
Multi-Tasks placement keeping	.40	.12	12.27	.000	1.49

Note. OR = Odds Ratio. Overall model χ^2 (7) = 45.96, p = .001.

3.10.4 The relationship between the Multi-Tasks measure and the Five Personality Subscales.

To determine whether personality traits were associated with the Multi-Tasks subscales, two linear regression procedures were conducted. In the first procedure, the Multi-Tasks Word Recall scores were regressed on the five personality traits. In the second procedure, the Multi-Tasks Placement Keeping scores were regressed on the five personality traits. The findings revealed that none of the personality traits were significantly associated with either of the two Multi-Tasks subscale scores. These are outlined in Appendix 2F.

3.11 Discussion

The main aim of the current study was to extend the findings of the first study in this series, exploring the structure, reliability and predictive validity of a new measure of cognitive ability, the Multi-Tasks test. Specific aims of the study were to: (1) determine if the Multi-Tasks test can be defined as a two factor model, similar to that shown in Study 1, but excluding a speed factor; (2) based on an extended protocol of cognitive ability measures, determine if the Multi-Tasks test measures three of the broad factors described by the CHC Theory of Intelligence, namely fluid Intelligence (Gf), crystallized intelligence (Gc), and short-term Memory (Gsm); (3) compare the predictive validity of the Multi-Tasks test with that of the single tests measuring Gf, Gc and Gsm in predicting job performance as measured by a four job performance indicators; (4) identify whether personality has an impact on performance on the Multi-Tasks tests; and (5) compare the predictive validity of the Multi-Tasks test with that of the personality measure in predicting job performance. Group demographic difference were also explored; whether the Multi-task tests varied by gender, level of education or age and whether the three versions behaved similarly in the analyses. These outcomes of the analyses of aims will be discussed in the sections that follow.

Supporting the findings of Study 1, the underlying structure of the Multi-Tasks test was found to consist of two main factors, Word Recall and Placement Keeping. Replicating Study 1, Broad Speediness (or speed of response) was dropped from the analysis and not included in the model of intelligence because it did not have a clear relationship with the other measures and a model incorporating speed did not fit the data well. Analyses revealed that while the Multi-Tasks test was positively correlated with all the cognitive ability measures, the highest correlation was with Digit Span. The model that best fit the data had Multi-Tasks and Digit Span loading on to Gsm and all other cognitive ability measures loading on to Gf. The model incorporating Gc did not adequately represent the data.

This outcome was slightly different to the findings of Study 1, as it appeared that introducing a greater range of cognitive ability measures further defined the place of Multi-Tasks within the structure of intelligence. Supporting the work of Stankov (1988), Multi-Tasks loaded onto Gsm, as did Digit Span, which has traditionally been used as a measure of Gsm. These tasks both involve holding information in short-term memory. Matrices and Digit Sequence involve non-verbal reasoning and have been widely used as measures of Gf. It was somewhat surprising that the Word Reasoning and Reading Comprehension tests loaded on to Gf , as these are traditionally used as measures of Gc. It appears that within the competing task paradigm, in this study at least, the tasks were solved using fluid intelligence rather than crystallized intelligence. This may be because these tasks, while relying on verbal skills, also have a reasoning component. Reading Comprehension also places demands on WM as it requires retention of verbal information. Thus, these tasks are perhaps more complex Gc tasks involving verbal reasoning and conceptual thinking. As shown in a range of studies, including Roberts et al. (1988; 1991), Carpenter et al. (1990) and Stankov (1999, 2003), the more complex the task, the greater the relationship with Gf.

Group differences analyses revealed only an overall effect for Education level on the Placement Keeping component of Multi-Tasks, yet post-hoc tests revealed no further group differences. There were no significant differences based on gender or age. This is in contrast to Study 1, which found significant effects for age, whereby older applicants did better than the younger applicants, but no effect for education. This may be due to the characteristics of the samples in each study, which will be further discussed later in the discussion.

Further, as in Study 1, time constraints need to be considered. The number of tests that were able to be administered was limited by the tolerance of the participants and the organisations involved. Testing sessions needed to adhere to a 70 minute timeframe.

The reliability coefficients for the Multi-Tasks measures were generally high and higher than those of the other cognitive ability measures. Overall, the Multi-Tasks tests therefore had greater reliability than the other measures, replicating the findings of Study 1.

Linear regression analyses revealed that none of the cognitive ability measures predicted Job Performance as indicated by the Job Performance Score (based on a rating system against unit or team objectives). In terms of Job Level, Placement Keeping consistently predicted Job Level whereas Word Recall did not. In the regression analysis with the other cognitive ability measures Word Recall had a significant and negative relationship with Job Level, and in the analysis involving the personality measures the relationship was non-significant. It is likely that for this sample, the Word Recall task was comparatively easy and therefore those at lower job levels focused more on this component to ensure that they answered as many correctly as possible. In contrast, more capable and senior level candidates may have tended to focus equally on both Word Recall and Placement Keeping, thus not performing as well as lower level applicants on Word Recall. Placement Keeping appears to be the better predictor of Job Level, likely because it was the more difficult task. The Multi-Tasks sub-tasks were both highly correlated with Years of Service, but did not significantly predict Years of Service when other measures were controlled for. While it might be expected that years of service would be confounded with age, this was found to be non-significant ($r = -.04$, $p > .05$). There was also no correlation between age and job level ($r = -.06$, $p > .05$). The participants appear to have been hired and promoted on factors mostly unrelated to age. In terms of the other cognitive ability measures, Matrices also predicted Job Level and Digit Sequence and Digit Span were positively correlated with Years of Service. Overall, Multi-Tasks was a stronger and more consistent predictor of more indicators of job performance than the other cognitive ability measures.

The relationship between Multi-Tasks and Job Level supports research indicating that competing task paradigms are particularly useful for predicting performance in complex and managerial jobs (Ben-Shakhar & Sheffer, 2001; North & Gopher, 1976; Sauer et al., 1999; Stankov et al., 1989). As jobs become less defined and more demanding, complex, novel problem solving tests appear to be the best predictors of job performance (Bertua et al., 2005).

There is some similarity between these findings and those of Fogarty and Stankov (1982) and Stankov et al. (1989) who used single measures of Gc (word meaning tasks), Gf (completing number series) and Gsm (digit span) and compared these with measures comprising subsets of these tasks combined into a competing task paradigm to predict ‘managerial potential’ of Australian employees. It was found that single tasks did not significantly predict managerial potential (as indicated by performance ratings, position and reporting steps to the Chief Executive Officer), however when the tasks were combined the predictive power became significant (Stankov et al., 1989). It was determined that the tasks measured the same underlying abilities, however there was a separate additional ability that only emerged on the competing tasks measures, that was predictive of managerial potential (Fogarty & Stankov, 1982). While the current study did not combine the individual intelligence measures to form a competing task measure but employed a new competing task paradigm, the outcomes regarding the predictive power of these measures were similar. In both studies the competing task measures were more complex, had a higher WM load and increased need to direct attention, compared to the individual tasks (Fogarty & Stankov, 1982; Stankov et al., 1989). It is these skills that are likely to be most valuable in complex jobs and roles involving managing others.

The other cognitive ability tests that predicted some aspects of job performance were Matrices, Digit Span and Reading Comprehension. Matrices is similar in kind to the Ravens

Progressive Matrices (RPM), a well-known task that has been found to measure deductive reasoning or ‘analytical’ intelligence, the ability to reason and problem solve using new information and novel concepts (Carpenter et al., 1990). It is also thought to be a good measure of g or general ability, particularly the non-verbal component of g (Gf) (Raven, Raven, & Court, 1998). Numerous studies have shown that scores on RMP are positively associated with performance in decision making roles and jobs requiring high levels of general mental ability (Fay & Frese, 2001; Gonzalez, Thomas, & Vanvukov, 2005; Raven, et al., 1998). Like Multi-Tasks, Digit Span was a Gsm measure, and it appears that holding information in mind was a necessary skill associated with better overall job performance in this population. As discussed in Study 1, tasks that assess verbal ability (such as Reading Comprehension) are likely to be important in job selection as the vast majority of jobs involve various forms of verbal communication, including written reports, email communication, meeting participation and presentations. In contrast to Study 1, Word Reasoning did not predict job performance in the current study and overall the relationship between job performance and cognitive ability measures other than Multi-Tasks was not as strong. There are certain characteristics of the participant population, the testing environment and the job performance measures which may explain these different findings, which will be discussed later. This outcome also highlights the importance of including another measure to more thoroughly assess the relationship between Gc and job performance.

Analyses revealed that none of the personality measures were positively associated with the Multi-Tasks test. It was hypothesised that there would be a relationship between performance on Multi-Tasks and Extraversion, given research indicating that Extraverts perform better on complex tasks placing high demands on WM. Surprisingly, however, given the correlation of the Multi-Tasks Placement Keeping and Word Recall with Gsm, there was no significant relationship between either of these components of the Multi-Tasks test and Extraversion.

While the Extraversion variable was normally distributed and there was a good spread of responses, it may be that other factors such as the demographics of the sample that influenced the results. There were almost twice as many female as male participants and approximately 68% were aged 29 years or less. A more varied group of participants may have been required to show an effect of personality on Multi-Tasks performance. Given the strong relationship between Gsm and Multi-Tasks, it may be that Multi-Tasks is more strongly associated with the short-term memory store component of WM rather than the executive component. As discussed, it is the executive component of WM that is associated with the DLPFC and ARAS system and found to be associated with better performance on such tasks by Extraverts, not the short-term memory component (Lieberman, 2000).

This is one of the pioneering series of studies examining the relationship between personality and Multi-Tasks, and it is recommended that further research be undertaken in this area.

Linear Regression with personality measures and the Multi-Tasks subscales revealed Multi-Tasks to be a better predictor of job performance than personality. None of the personality measures predicted any indicator of job performance, whereas at least one of the Multi-Tasks sub-tasks predicted Employment Status and Job Level and both predicted the overall indicator of objective job performance. Given that none of the other cognitive ability measures were found to be consistent and strong predictors of job performance, it is unsurprising that personality failed to predict job performance. Research consistently shows cognition to be the most reliable predictor of job performance (Bertua, et al., 2005; Gottfredson, 1986, 1997; Hunter, et al., 1990; O'Reilly & Chatman, 1994; Salgado, et al., 2003; Schmidt & Hunter, 1992; Schmitt & Mills, 2001). Further, there is research which shows that personality is a better predictor of subjective or indirect indicators of job performance, such as effort, perseverence, leadership and personal grooming, while cognitive ability measures are better indicators of job-specific and task proficiency (McHenry, Hough, Toquam, Hanson & Ashworth, 1990). The personality inventory is based on the very well researched Five Factor Model of Personality and the correlations in the present study are consistent with previous studies, particularly those for Extraversion, Neuroticism and Conscientiousness (Costa, McCrae, & Dye, 1991; Digman, 1997), therefore the results are not believed to be reflective of issues with the measure.

The results were limited by the job performance indicators that were available. While a range of measures were collected, job performance measures were not available for all participants and the job performance scores were not consistent between participants. More thorough job performance data may have allowed for a stronger relationship between job performance and Multi-Tasks to be revealed, as per Study 1. While the lack of consistency in job performance data appeared to make it more difficult to obtain significant results, the advantage is that the efficacy of Multi-Tasks is demonstrated within a naturalistic work setting with the challenges that this entails. Well controlled research within laboratory settings may not generalise well to the workplace, whereas this series of studies is directly relevant to work settings.

The results of the current study and Study 1 into the effectiveness of the Multi-Tasks test have indicated that there are extensive opportunities to expand the scope of this research. The current findings suggest that the Multi-task measure is able to predict intelligence and job performance over and above single psychometric tests and personality measures, both of which are commonly used in job selection. This may be due to the engaging and complex nature of the task.

There are a number of limitations to the study and unique features of the participants, variables and testing environment that need to be considered when interpreting the results. As previously discussed, one limitation is that the participants were not representative of the working public. There were twice as many females as males, approximately 68% were aged 29 or less and approximately 43% were 24 years of age or less. Approximately 89% of participants were voluntary transfers at Job Level 1 and the majority had less than 1 year of service with the organisation. Stronger relationships between cognitive ability measures, personality and job performance may have been established had a wider group of individuals been included in the study. For instance, a large meta-analysis has shown that male adults and adolescents (over the age of 15 years) perform significantly better than females on several versions of the RMP (Lynn & Irwing, 2004). Given the high proportion of females in the current sample, this may have contributed to the relatively low scores and low reliability of the Matrices test and possibly its lack of relationship with most indicators of job performance. Given that the Multi-Tasks test appears to be strongly predictive of performance in complex, managerial roles, stronger relationships between Multi-Tasks and job performance may have been revealed with older, more experienced individuals. Study 1 found significant effects of age and gender on the Multi-Tasks test, with older participants performing better on the Word Recall component and males performing better on the Placement Keeping component. Further, Study 1 included a similar number of males (54%) and females and more than 60% were aged 30 years or older.

It is also important to note that a number of the cognitive ability measures and the personality measures were highly skewed, with many participants scoring at or close to ceiling. This was the case with the Multi-Tasks measure, the Gsm measure (Digit Span) and the Conscientiousness and Agreeableness components of the personality scale. In relation to the cognitive ability measures, it is possible that the tasks were not complex enough for this population. This is unlikely however, as previous research has shown competing task methodologies to be complex (Carpenter et al., 1990; Roberts, Beh, Spilsbury & Stankov, 1991; Stankov, 1999; 2003). It is also possible that the high proportion of young people in the sample skewed the results, as it is known that performance on measures of Gf declines with age (Belsky, 1990; Bugg, Zook, DeLosh, Davalos & Davis, 2006; Horn & Noll, 1997; Moutafi et al., 2004). This is unlikely to completely explain the findings, as it would not explain the skew in the Gc measures (which are thought to either improve with age or remain stable) (Belsky, 1990; Kaufman, Johnson & Liu, 2008) or the personality measures. An alternative explanation, which would explain the skew on the Multi-Tasks test, Digit Span test and the personality measures, is that participants may have employed strategies to maximise their performance. In relation to the cognitive ability measures, the participants may have had the opportunity to make notes to aid recall. The Digit Span and Multi-Tasks test are both amendable to note-taking as they require keeping track of the order of presented items. The tasks were undertaken at the workplaces involved in the study, which were located throughout Australia, and were not able to be overseen by the author. While supervision instructions were provided to test invigilators (see Test Administration Instructions in Appendix 3.5) it is unclear if supervision was consistent between workplaces. It is possible that participants at some worksites may have taken advantage of lax supervision to try to maximise their scores. Despite the fact it would not be in the organisations' interests to allow candidates the opportunity to embellish their scores, it is a possible explanation for the skewed results for the Multi-Tasks test and Digit Span. This was not anticipated as in a pilot study involving 148 job applicants for roles similar to those in Study 2, the mean scores were much lower, 8.58 (SD=3.46) and 29.14 (SD=10.04) for Word Recall and Placement Keeping respectively, and the scores were normally distributed. Further, there was no evidence of embellishment in the results for both Multi-Tasks and the other cognitive ability measures used in Study 1. In terms of the personality measures, many studies including large meta-analytic studies have demonstrated that job applicants have the tendency to embellish their responses on personality scales, particularly those scales (e.g., Conscientiousness) where high scores are clearly more favourable (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Rosse, Stecher, Miller, & Levin, 1998; Viswesvaran & Ones, 1999). In Birkeland et al.'s (2006) meta-analysis of thirty-three studies comparing job applicant and non-applicant scores on the Big Five personality scales, applicants across all job types scored significantly higher on four of the five scales (O, C, E and N). The rank ordering of scales varied depending on the job type, such that applicants scored highest on scales thought to be directly relevant to the job they were applying for. It may be that in this population, applicants determined that Conscientiousness and Agreeableness were likely to be particularly relevant to job performance and answered accordingly.

This study extended Study 1 by including a number of job performance indicators, both objective (Employment Status, Job Level and Years of Service) and subjective measures based on ratings against job objectives. While the range of measures was useful, there were disadvantages to this approach. Study 1 utilised a template with clear rating guidelines that was provided by the author. In the present study, the client provided their own performance framework. The job performance scores were not consistent between participants as different business units (and even some teams within the business units) used a different score card with different ratings, and these were changed annually and sometimes quarterly. Further, job performance indicators were not available for all participants. The sample size for the subjective job performance measures was particularly low (n=50).

This may explain why Multi-Tasks, the other cognitive ability tests and the personality inventory did not predict any subjective measures of job performance. This is in contrast to Study 1 in which Multi-Tasks and the other cognitive ability measures predicted at least two of the three subjective measures of job performance. It is likely that those objective indicators of job performance that Multi-Tasks was predictive of are those that are most indicative of higher general mental ability (being at a higher job level or transferred due to good performance), thus supporting the research that Multi-task performance is correlated with intellectual functioning (Ben-Shakhar & Sheffer, 2001; Stankov, 1988, 1989, 2003).

Analyses revealed that the measures chosen as Gc measures (Word Reasoning and Reading Comprehension) were more strongly associated with Gf. They appeared to be complex verbal tasks and involved a reasoning component. To further investigate the place of Multi-Tasks within the CHC model it may be useful in future research with English speaking participants to use simpler verbal measures drawing more on store of acquired knowledge rather than reasoning, such as vocabulary or general knowledge tasks. For instance, the Gc measures in the Woodcock Johnson Tests of Cognitive Ability measure lexical knowledge, word meanings and general knowledge (McGrew & Woodcock, 2001). Only one component of one task (analogies) has a reasoning component.

This study addressed a number of the limitations raised in Study 1, including the need to further investigate Multi-Tasks within the CHC Theory and to compare its predictive validity against a wider range of cognitive ability measures and other well-known measures used in personnel selection, including personality assessment. One limitation not addressed is the issue of cross cultural comparisons. Further, given the reliability and predictive validity of the Multi-Tasks test, particularly for complex and managerial roles, it will be important to develop versions of the Multi-Tasks that are highly generalisable and culturally fair. Given its correlation with Gf, which is relatively free from education and cultural bias, it is likely that Multi-Tasks will be valid cross-culturally. Those respondents with English as a second language were not well represented in the data collected and therefore did not allow for proper investigation of the differences between groups that differ in their primary spoken language. As this test is computer-based, opportunities now exist to extend research quickly and easily to other countries. This would yield information about the importance of intelligence to job performance on an international level.

CHAPTER 4

STUDY 3

COMPETING TASKS AS MEASURES OF INTELLIGENCE AND PREDICTORS OF JOB PERFORMANCE

4.1 Summary of Study 1 and Study 2

The first study in this series introduced a new measure of cognitive ability, the Multi-Tasks test, which employed a competing tasks paradigm. Study 1 explored the relationship between the Multi-Tasks test and job performance as measured by supervisor ratings, and compared its efficacy to measures of general cognitive ability (fluid and crystallized intelligence) which form part of the Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities (McGrew, 2004; 2009). These have been widely used as predictors of job performance.

The second study extended the analysis to further define the Multi-Tasks test within the CHC Model of Cognitive Abilities and to compare the predictive validity of Multi-Tasks with a wider variety of cognitive ability measures, including additional measures of Gf and Gc and a measure of short-term memory (Gsm). Gsm is a limited capacity system which allows one to maintain awareness of events that have just occurred (e.g., the last minute or so). The usefulness of Multi-Tasks in predicting job performance was also compared with one of the most popular form of employment test, the personality inventory (Barrick, et al.; 2001; Hogan, 2005; Hogan & Holland, 2003).

These studies revealed the Multi-Tasks test to be a highly reliable measure, and more reliable than the other cognitive ability tests in Study 1 and Study 2 and the personality measure in Study 2. It was also a better predictor of job performance as measured by supervisor ratings (Complex Performance and Promotional Potential) in Study 1 and objective job performance indicators (Job Level and Employment Status) in Study 2.

The studies provided support for previous research on competing tasks: the Multi-Tasks test was found to be more strongly related to Gf than Gc (Carpenter, et al., 1990; Roberts, et al., 1988; Roberts, Beh, et al., 1991); had the strongest relationship with Gsm, (Stankov, 1988); and was particularly useful for predicting managerial potential and performance in complex jobs (Ben-Shakhar & Sheffer, 2001; North & Gopher, 1976; Sauer, Wastell, & Hockey, 1999; Stankov, 1982; Stankov et al., 1989). The studies also extended previous research by revealing the underlying factor structure of the Multi-Tasks measure, demonstrating the efficacy of Multi-Tasks in predicting various aspects of real world job performance, and revealing it to be a more effective measure than personality assessment in job selection.

Study 1 revealed that the Multi-Tasks measure comprised three factors, one for each of the tasks that it requires (keeping track of ball movements, referred to as placement keeping and remembering the order of words presented, or word recall) and a factor for speed of response. Study 1 also found that the other cognitive ability measures included a speed (or speediness) factor. As discussed in Study 2 the speed factors were not found to operate consistently or to assist with interpretation of the Multi-Tasks model or the model of intelligence. Further, the research on the relationship between speed and job performance, particularly in professional and complex roles was lacking. The speed factors were therefore removed from the analysis. This will also be the case in the present study.

4.2 The Multi-Tasks test and the CHC Theory of Cognitive Abilities

As demonstrated in Study 2, of most relevance to the competing task paradigm are the Gf, and Gsm factors (König, et al., 2005; Stankov, 1988; Stankov & Roberts, 1997). In the context of a competing task paradigm, which draws on limited processing resources to allow an individual to respond to two complex tasks within the same period of time, the ability to hold in mind information while it is being processed (Gsm) and to process the information using deliberate and controlled mental operations (Gf) are key cognitive requirements.

These abilities were also shown in Studies 1 and 2 to be associated with better performance at work. There is growing support for the usefulness of the Multi-Tasks test in predicting performance within the workplace. In addition to its usefulness as a measure of intellectual functioning, the ability to perform two tasks simultaneously is of value in the increasingly complex, fast paced world. Workplaces are becoming increasingly diverse due to globalisation and internationalisation of business. The ability of tests to be consistent across racial groups is extremely important in our increasingly multicultural society. It is therefore important to investigate the usefulness of this measure in diverse groups of people.

The current study will investigate the validity of this test cross-culturally, focusing on job applicants in China. The performance of participants from Chinese workplaces on a Chinese language version of the Multi-Tasks test will be examined. The relationship between performance on Multi-Tasks and job performance will also be explored, in order to begin to investigate whether the Multi-Tasks test is a valid predictor of job performance cross culturally.

The discussion commences with a review of the history of cognitive testing in China and its relationship with the West.

4.3 History of Cognitive Testing in China

Cognitive ability testing has a long and interesting history in China. Ancient Chinese philosophers put forward assessment criteria for intelligence based on education, management, and social activities (Higgins & Xiang, 2009). Intelligent people were those who could make inferences and predictions based on available information and who had insight into themselves and other people. The political thinker and educator Confucius classified his students into three categories: ‘people of great intelligence’, ‘people of average intelligence’ and ‘people of little intelligence’ (Higgins & Xiang; Jing, 1994), which as a basic classification system is still in place today. By the 7th century AD the Chinese system of public service exams was developed, which assessed candidates’ ability to remember and interpret Confusion texts and assigned official positions based on performance on these exams (Chan, Shum, & Cheung, 2003; Higgins & Xiang, 2009). This idea was later adopted by the British Government (Higgins & Xiang, 2009).

In the early 20th Century the intellectual reform movement in China and increased contact with Western societies facilitated the development of local cognitive ability tests and the translation and revision of Western measures (Higgins & Xiang, 2009; Jing, 1994). These included the 1924 Chinese version of the Binet-Simon test, widely regarded as the first psychological test of cognitive processes (Chan et al., 2003; Higgins & Xiang). Following the founding of The People's Republic of China in 1949, and particularly during Chairman Mao's Cultural Revolution in the late 1960's and early 1970's, cognitive testing virtually ceased. The Marxist doctrines of the Communist Party viewed psychology as hypothetical pseudo-science and the study of individual differences as 'bourgeoisie' (Higgins & Xiang, 2009; Jing, 1994). They also sought independence from Western ideas (Jing, 1994). Psychology departments at universities were closed down and academics were sent to work in labouring jobs.

From 1978 the field was rehabilitated and cognitive ability testing flourished as psychologists gained greater understanding of neuroscience, neuropsychology and its clinical applications and China launched a policy of reform and openness to the outside world (Chan et al., 2003; Jing, 1994). Because of the long period of inaction the most efficient means of renewing intelligence research was to adapt cognitive tests developed in other (mainly Western) countries (Chan et al., 2003). Many popular tests were adopted by Chinese psychologists and psychiatrists in the 1980s and 1990s, including the Weschler series of adult and child intelligence and memory tests and Raven's Progressive Matrices (Higgins & Xiang, 2009). By the 1990's indigenous tests were also being developed, including Gong's test of non-verbal intelligence (for ethnic minorities) (Gong et al., 1995) and Group intelligence test for Army Academy students (Liu et al., 2001).

While in the last 30 years China has become more open to the outside world and receptive to Western ideas, it remains a collectivist culture with a significantly different cognitive orientation to Western individualistic societies. China's 'holistic' cognitive orientation emphasises relationships and connectedness (Ji, Peng & Nisbett, 2000). Conversely, Western cultures such as North America and Australia have an 'analytic' cognitive style, which focuses on individual differences (Ji et al., 2000).

4.3.1. Use of cognitive ability tests cross culturally.

The pace with which foreign cognitive tests were adapted in China following the end of the Cultural Revolution raises the issue of cross-cultural validity. Language, education and social/cultural differences can influence performance on cognitive tests, particularly when two cultures are quite disparate. Despite this, tests such as the Weschler series have been translated into various languages and used in other cultures successfully, and efforts have been made to develop 'culture-fair' assessments. The challenges associated with using tests adapted from foreign cultures and examples of successful adaptation of tests cross-culturally will be discussed in the following sections.

4.3.2. Challenges associated with using foreign ability tests.

Chan et al. (2003) reviewed 123 studies that described cognitive assessment instruments used in Asia (primarily Japan, China and India) between 1981 and 2001. The studies included a total of 36 assessment instruments. The vast majority of instruments were adapted from foreign tests rather than being locally developed (30 compared to 6). Only 8 of the instruments, 3 of which were locally developed, fulfilled Chan et al's (2003) criteria for clinically validated tools, being: validity tested, local normative data with a sample size of greater than 50, and for adapted tests, translation and cross-checking of translation and empirical cross-cultural comparison to demonstrate applicability to the local population. A number of tests met all criteria other than cross-cultural comparison (Chan et al., 2003). The influence of culture on test performance was demonstrated using the example of verbal fluency tasks (which involve naming as many items in a category or beginning with a certain letter as possible in 60 seconds). Experience of and exposure to certain categories (e.g., items in a supermarket) affects performance on these tasks (Chan et al., 2003). The authors discussed the importance of using clinically validated, preferably locally developed, tools that best suit the cultural and linguistic characteristics of the local population.

Researchers have outlined many reasons for questioning the cross-cultural validity of ability tests. These include differences in educational systems, cultural values, social systems and language (Campbell & Xue, 2001; Chan et al., 2003; Fan, Zhang, & Watkins, 2010). An emerging area of research is differences in the neurocognitive processing of cultural groups. For example, it has been shown that Chinese language processing uses more brain regions than English language processing, which tends to be lateralised to one (predominantly the left) hemisphere (Chan et al., 2002, as cited in Chan et al., 2003). It is also known that bilingualism has an impact on neurocognitive processing (Campbell & Xue, 2001). Assessment tools based on neurocognitive models may therefore not be generalisable to other cultural groups (Chan et al., 2003).

Unlike individualistic societies such as North America and the United Kingdom (where most cognitive ability tests have been developed), collectivist cultures such as China emphasise conformity and obedience, and schools and universities encourage and reward compliance, not creativity (Fan et al., 2010; Varela, Salgado, & Lasio, 2010). Collectivist organisations and systems encourage cooperative behaviours and social integration while individualistic ones reward and promote individual achievement and competition (Varela et al., 2010).

The generalisability of Perry's (1970) theory of intellectual and ethical development, which has been used to assess the development of young adults' reasoning ability in the U.S., to Chinese students was studied by Zhang (1999) using the Zhang Cognitive Development Inventory (ZCDI) which is based on Perry's theory. 1311 Chinese students from 12 universities in China (820 from Beijing and 503 from Nanjing) and 152 U.S. college students responded to the ZCDI. The reliability coefficients and validity statistics were found to be acceptable for both Chinese and U.S. groups. However, the Chinese groups, particularly those from Beijing, demonstrated a cognitive-developmental pattern that was reversed from the one proposed by Perry and which was demonstrated by the U.S. group. Chinese students from lower level college classes demonstrated significantly higher reasoning on Perry's scale than those from higher levels (Zhang, 1999). The differences were explained by the differing educational systems. Chinese students had little choice in the classes they undertook as part of a course, were assigned to a certain class section when they enrolled which remained constant throughout their studies, were trained in very narrow fields of specialisation, and assigned to a certain residential hall (Zhang, 1999). This severely restricted their ability to undertake independent decision-making and reasoning, compared to the U.S. students who had a significant degree of choice in the structure of their courses and who they interacted and lived with (Zhang, 1999). Since this study was published it is understood that universities have undergone significant reform and have introduced more freedom of choice, however the changes have been slow and the government continue to maintain a considerable degree of control over higher education policy (Mok, 2005).

Asian students tend to outperform other students in certain areas, particularly in mathematics. Asian students in the U.S. have be shown to have higher mathematics SAT scores than Caucasian and African American students. In the period from 1987 to 1996 the scores for Asian students were 37 points higher than Caucasian and 144 points higher than African American students (U.S Department of Education, 1997). East Asian educational systems show a greater concentration on mathematics instruction and place more importance on mathematics achievement than North American systems (Campbell & Xue, 2001).

A variety of informal cultural differences have also been shown to affect learning and academic outcomes, which also influence outcomes on intelligence testing. These include parental pressure and high expectations of performance and pursuing extra-curricular activities to accelerate learning (Campbell & Xue, 2001).

Finally, unlike Western cultures, when reporting on intellectual outcomes, Asian researchers tend to focus on environmental influences and situational or contextual differences, rather than individual differences (Ji et al., 2000). This is because of the communist ideology and their recent history of being an agricultural economy in which cooperation and obedience were required (Higgins & Xiang, 2009; Ji et al., 2000).

4.3.3. Successful adaptation of measures and 'culture-fair' assessments.

Despite the cultural and language differences, a number of cognitive assessment tools developed in foreign countries have been successfully adopted in East Asian cultures. Five adapted tests in Chan et al.'s (2003) study (primarily dementia assessment tools) were found to be valid and reliable measures. These tests, developed in the U.S. and Europe, were sensitive to the domain they claimed to measure, differentiated between groups (e.g., controls and brain injured people) and had factor structures that revealed theoretically and clinically meaningful underlying processes for people in China, Hong Kong, and Korea (Chan et al., 2003).

The validity of a translated version of the U.S. developed Thurston Primary Mental Ability Scale, which measures five aspects of intellectual functioning (spatial relations, inductive reasoning, verbal ability, word fluency and number skills) was investigated using a large sample of 413 Hong Kong Chinese adults. The scale was found to have a high internal consistency (.9), and moderate to high convergent validity (.4-.76) when compared with another measure of intellectual functioning (Chou & Chi, 2000).

A number of assessment tools have been designed to transcend cultural differences. Ravens Progressive Matrices for instance, a tool that assesses fluid reasoning, has been shown to have similar psychometric properties across various groups, including African Americans, Caucasians, Hispanic, Asian, African and East Indian groups (Rushton, Skuy & Bon, 2004; Strauss, Sherman & Spreen, 2006). It has a reasonably low correlation with academic achievement and is considered a ‘culture fair’ ability measure (Strauss et al., 2006).

4.3.4. Performance of Chinese participants with specific cognitive domains.

In addition to studies investigating broad intelligence measures in Asian cultures, an increasing number of studies have also explored performance on tasks within specific cognitive domains, and compared this with outcomes in Western cultures. Those of particular relevance to the competing task paradigm are discussed in the following paragraphs.

4.3.4.1 Working memory.

WM is a system whereby a controlling attentional system (the central executive) coordinates two ‘slave’ systems, the phonological loop (which processes verbal information) and the visuo-spatial sketchpad (which processes non-verbal information) to store and manipulate information, often in the presence of interference (Baddeley, 1990; Luer et al., 1998; Oberaur & Gothe, 2006).

Baddeley’s model demonstrates the importance of language on working memory performance. The phonological loop is comprised of a temporary store of speech related information which fades quickly, and an articulatory control process whereby the information in the store is refreshed through rehearsal (e.g. reciting a telephone number over and over to allow it to be held in mind until it is no longer needed) (Luer et al., 1998). Visual information can also be manipulated in this way by converting it to language based information. The articulation speed of words is an important determinant of how well the phonological loop, and thus the working memory system, performs. The shorter the articulation speed, the more items that can be held in short-term memory store and the larger the memory span (Lass, 1997, as cited in Luer et al., 1998).

In a German language paper, Lass (1997, as cited in Luer et al., 1998) conducted an immediate recall and item recognition study involving 192 students, half of whom were Chinese. They found that the Chinese students had significantly larger memory spans than the German students; however they did not explore the reasons for this. In their study, Luer et al. (1998) replicated these results. They also found that Chinese participants had higher intelligence scores (on a measure of fluid intelligence). In a series of experiments, the authors demonstrated that the memory span effect remained when intellectual functioning was controlled statistically, and the difference disappeared in tasks where participants could not use the phonological loop to refresh items in memory. They traced the differences back to shorter articulation times by the Chinese as compared to German participants (Luer et al., 1998). It appears that the Chinese language facilitated more efficient use of the phonological loop and better immediate memory performance.

Luer et al.(1998) discussed their results in the context of Cattell’s (1971) theory of crystallized (Gc) and fluid intelligence (Gf). They advised that while Gf is usually assumed to be a culture independent measure of intelligence, their study demonstrates the impact of cultural differences on working memory, which is an underlying element of Gf.

4.3.4.2 Attention.

Attention, particularly the ability to effectively control and manage attention, has been shown to be an important element in competing task performance (Kahneman, 1973; Lorsbach & Simpson, 1988; Goonetilleke & Luximon, 2010). Competing tasks are often used to examine the relationship between intelligence and/or performance with age. An age-related decrement in performance on competing tasks and other tasks drawing on Gf has been found for older adults in Western cultures. This has been hypothesised to reflect attention differences between younger and older participants, particularly the capacity to divide and maintain attention (Lorsbach & Simpson, 1988).

Zhou, Fan, Lee, Wang, and Wang (2011) found a similar pattern of results in their study using a sample of 90 Chinese participants in three age groups, young (20-38 years), middle aged (40-59 years) and older (61-80 years) adults. They found a clear age effect on the higher level executive (self-regulation of cognition) and alerting (maintaining alertness) components of attention, but not the orienting (selection of information from sensory input) component (Zhou et al., 2011). The decline was proposed to be associated with deterioration in the neural networks of the pre-frontal cortex which are responsible for executive functioning and deteriorate more rapidly than other cortical regions (Resnick et al., 2000; 2003). The pre-frontal cortex has also been found to be activated during competing task performance (Newman, Keller & Just, 2007; Szameitat, Schubert, Müller, & von Cramen, 2002). The authors refer briefly to the impact of aging on the ability to 'multi-task' (or perform different tasks simultaneously), which is associated with the decline in executive control of attention (Zhou et al., 2011).

These individuals were recruited from one province in China, the Anhui Province which is a fairly poor, largely agricultural region. As will be shown in the next section, there are significant differences between the cognitive abilities of various groups within China, and the results of the study therefore may not generalise to other parts of China. For this reason, further research is needed. This study will examine the cognitive abilities of two groups, Chinese nationals and people of Chinese origin working in Australia to further illuminate the impact of culture on competing tasks performance.

4.4 Group Differences in Cognitive Ability within China

There are a number of group differences in cognitive ability within the Chinese culture which are worthy of note. Perhaps the most interesting is the impact of the single child policy on intelligence and achievement outcomes.

4.4.1. 'Single' children.

Since the single child policy became law in 1979/1980 an interesting area of difference unique to China has emerged: the child in the enforced single child family (Jing, 1994). Comparisons of single child families can be made with rural families (where the policy was not so strongly implemented), ethnic minorities and studies conducted before the one child policy was implemented. Single children have consistently been found to have higher intelligent quotients (IQ) than children with siblings. In certain 'key' high performing schools the average IQ score was found by Wang, Xue and Li (as cited in Higgins & Xiang, 2009) to be 124.3, with 41% of children scoring above 130 and 81% above 120. In Western cultures an IQ of 120 is at the 91st percentile and an IQ of 130 is at the 98th percentile, meaning that approximately 9% and 2% of individuals would be expected to function at these levels respectively. The explanation given for this outcome is that parents of single children appear to give them better opportunities than other parents to maximise intellectual development (Higgins & Xiang, 2009).

4.4.2. City vs. rural.

Individuals living in cities have been found to have higher IQ's than those living in rural areas in China (Higgins & Xiang, 2009). This difference is most apparent in early childhood and has been attributed to city children having access to formal education at an earlier age, a higher quality of education, better access to resources and better educated parents. There even appear to be differences between remote and less remote rural areas (Higgins & Xiang, 2009).

4.4.3. Gender differences.

Some studies on intellectual differences have shown a slight IQ advantage for boys (see Higgins & Xiang, 2009 for a summary of Chinese language studies). The differences tend not to be significant however and are most apparent in early childhood. It may be that the one child policy (which has resulted in boys outnumbering girls) has influenced these results.

4.5 Intelligence and Later Success

In Western societies, intelligence has been shown to be highly significantly correlated with occupational outcomes including better performing ratings, and higher occupational and income level (Bertua, et al., 2005; Gottfredson, 1986, 1997; Judge, Higgins, Thoresen, & Barrick, 1999; Salgado et al., 2003; Schmitt & Mills, 2001; Wilk, Desmarias, & Sackett, 1995). In China intelligence testing tends to be largely confined to clinical and educational settings. Studies have demonstrated a strong correlation between high IQ and later outcomes including academic achievement and better educational opportunities (e.g. university entrance) (Higgins & Xiang, 2009). There have been few studies on the relationship between intelligence and work performance. It is likely however that if intelligence is associated with better academic outcomes and educational opportunities, it will also be associated with better job opportunities and work performance.

4.6 Study Aims

The main aim of the current study was to explore the cross-cultural validity of the Multi-Tasks test and its efficacy in predicting job performance within the Chinese culture. Aims of the study were to: (1) determine the effect of language and culture on Multi-Tasks by examining the performance of Chinese workers within the Chinese workforce, (2) determine whether the Multi-Tasks test predicts job performance within a Chinese context, (3) determine if Multi-Tasks replicates a three factor theoretical model and its relationship with measures of Gf, Gc and Gsm, and (4) explore demographic differences, specifically whether performance varies by age, gender, job level, and level of education.

As demonstrated in Study 1 and Study 2 Multi-Tasks correlates highly with Gf and Gsm, and therefore, it was hypothesised that performance would not be affected by language or culture. It was hypothesised that the Chinese workers would perform similarly to English speakers in the Australian workforce. It was further hypothesised that, replicating the findings of Study 2, the underlying theoretical model would be a three-factor model. Multi-task performance was hypothesised to predict job performance, and more experienced and higher level participants would perform better than less experienced and lower level participants.

Three models were proposed to explore Multi-Tasks and its place within the CHC model of intellectual functioning. Figure 3.1 outlines the proposed model for the underlying structure of the competing tasks measure (Multi-Tasks). Based on the results of Study and Study 22 it was hypothesised that the underlying dimensions of the competing task measure would include factors relating to each of the competing tasks (word recall and placement keeping). As with Study 2, the speed component of the measures is being excluded from this study.

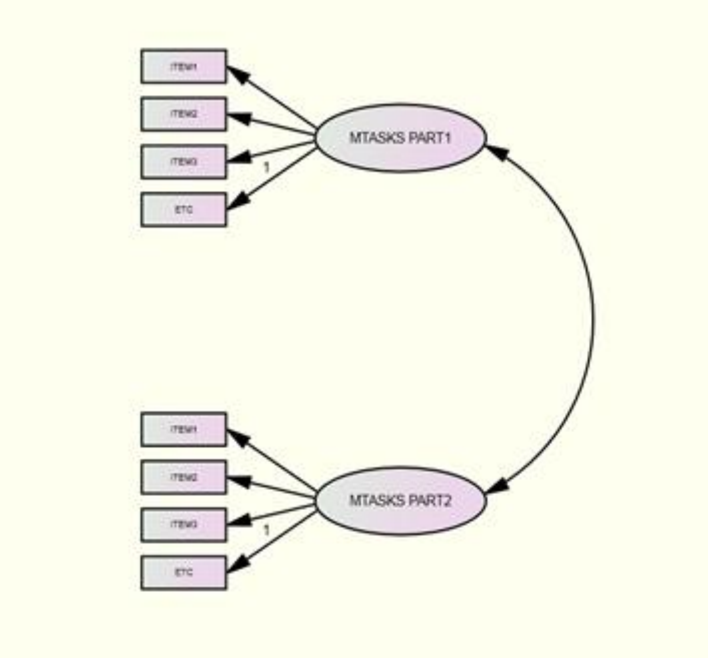


Figure 3.1. Proposed measurement model for Multi-Tasks.

Figures 3.2 and 3.3 outline the hypothesised relationships between the competing tasks measure and other well-known measures used in job selection and the CHC theory of intelligence. Based on the outcomes of Study 1 and previous research (e.g., König, et al., 2005; McGrew, 2009; McGrew & Woodcock, 2001; Stankov, 1988) it was hypothesised that the components of the Multi-Tasks measure would be strongly associated with Gf, as would the Digit Sequence, Number Series, and Matrices tasks, whereas the Word Reasoning and Reading Comprehension tasks would load onto the Gc factor. Digit Sequence, Number Series, and Matrices are novel non-verbal problem solving tasks, while Word Reasoning and Reading Comprehension are verbal language reasoning and comprehension tasks. This is shown in Figure 3.2.

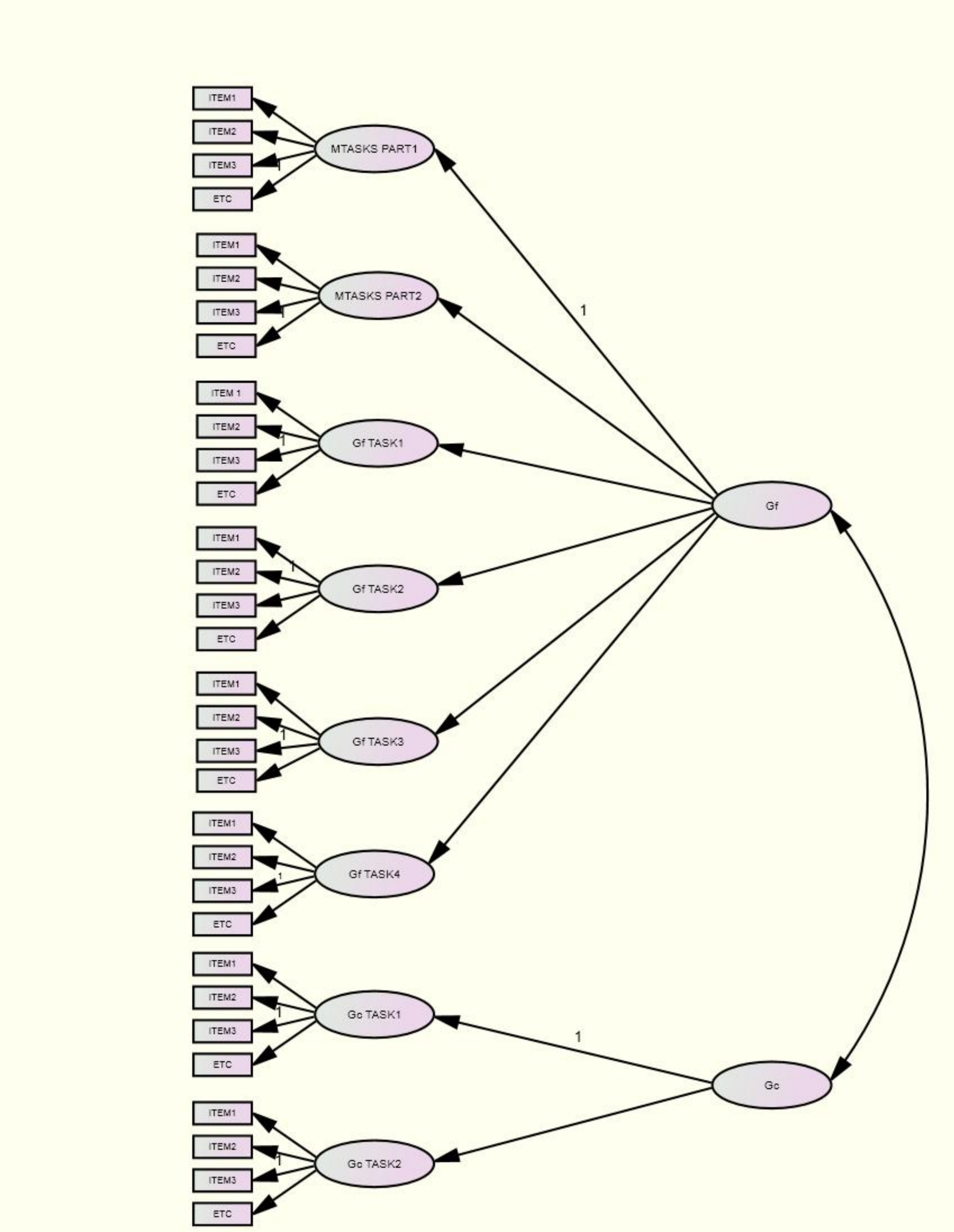


Figure 3.2. Proposed measurement model for Multi-Tasks and other intelligence measures.

Introducing Gsm into the model was proposed to further define the place of Multi-Tasks within the CHC theory. It was hypothesised that because competing tasks have been shown to have a higher correlation with Gsm than single tasks (Stankov, 1988) and due to the high demands of the Multi-Tasks test on WM which is highly correlated with Gsm (McGrew, 2004), Multi-Tasks would have a strong relationship with Gsm, as would Digit Span, which is a Gsm task (McGrew, 2009: McGrew & Woodcock, 2001). Matrices and Digit Sequence would continue to load onto the Gf factor and Word Reasoning and Reading Comprehension on to the Gc factor. This is represented in Figure 3.3. Speed was removed from the model in Figure 3.3 as Study 1 revealed and Study 2 supported, that a model incorporating speed measures did not fit the data well. It was hypothesised that this would also be the case in the current study.

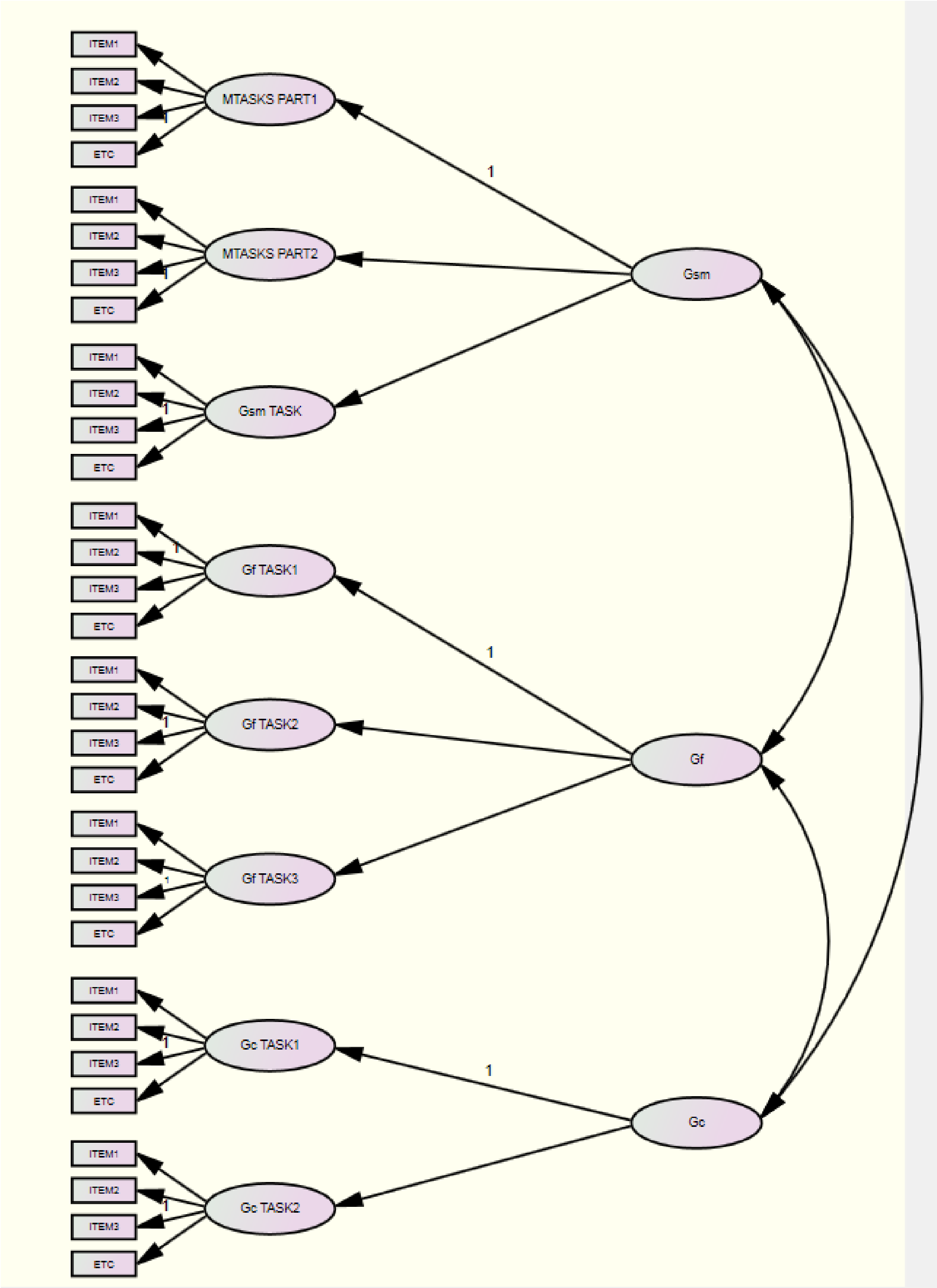


Figure 3.3. Proposed measurement model for Multi-Tasks and other intelligence measures (2).

4.7 Method

4.7.1 Participants.

A total of 101 job applicants for a range of positions within a manufacturing organisation in China participated in the study. The group consisted of 73 (72.3%) males and 28 (27.7 %) females. The age distribution of the participants is displayed in Table 3.1 and their levels of education are shown in Table 3.2.

Table 3.1

Age of Participants (N =101)

Variable	Frequency	Percentage
Age		
18 to 24	16	15.8
25 to 29	30	29.7
30 to 34	20	19.8
35 to 39	20	19.8
40 to 44	9	8.9
45 to 49	4	4.0
50 to 54	2	2.0

Table 3.2

Education Level of Participants (N = 101)

Education Level	Frequency	Percentage
Middle school diploma	43	42.6
High school diploma	24	23.8
Technical college	16	15.8
Took academic courses	4	4.0
Bachelor degree	9	8.9
Masters degree	5	5.0

4.7.2 Materials.

The study employed seven measures of cognitive ability and three measures of job performance. Six of the ability measures were the same as those used in Study 2, being: Digit Sequence and Matrices (source: E-ntelligent Testing Products Pty Ltd, criterion measures of fluid intelligence), Word Reasoning and Reading Comprehension (source: E-ntelligent Testing Products Pty Ltd, criterion measures of crystallized intelligence), Digit Span (source: Fogarty, unpublished), a measure of short-term memory) and the Multi-Tasks test (source: Bradley Dolph). An additional numerical type measure of fluid intelligence was also used, Number Series (E-ntelligent Testing Products Pty Ltd) as the reliability of Digit Sequence was found to be relatively low in the previous studies. All measures were translated into Chinese by two translators who were accredited by the National Accreditation Authority for Translators and Interpreters. One translator translated the items from English into Chinese and the other checked the translations for accuracy. In Study 2, Word Reasoning and Reading Comprehension had greater loadings on the fluid intelligence factor while in Study 1 Word Reasoning loaded onto the crystallized intelligence factor. Models with both factor loadings will be tested in this study.

The measures of job performance consisted of three objective indicators of job performance, Absenteeism, Unreliability and Job Level. Each of these tasks will now be described in more detail.

4.7.2.1 Digit Sequence test.

Digit Sequence was also used as a Gf measure in Study 3. It was administered in the same manner as the previous studies yet the items were translated into Chinese. For a full description of the task refer to Study 1. Figure 3.4 provides an illustration of a Digit Sequence item in Chinese.

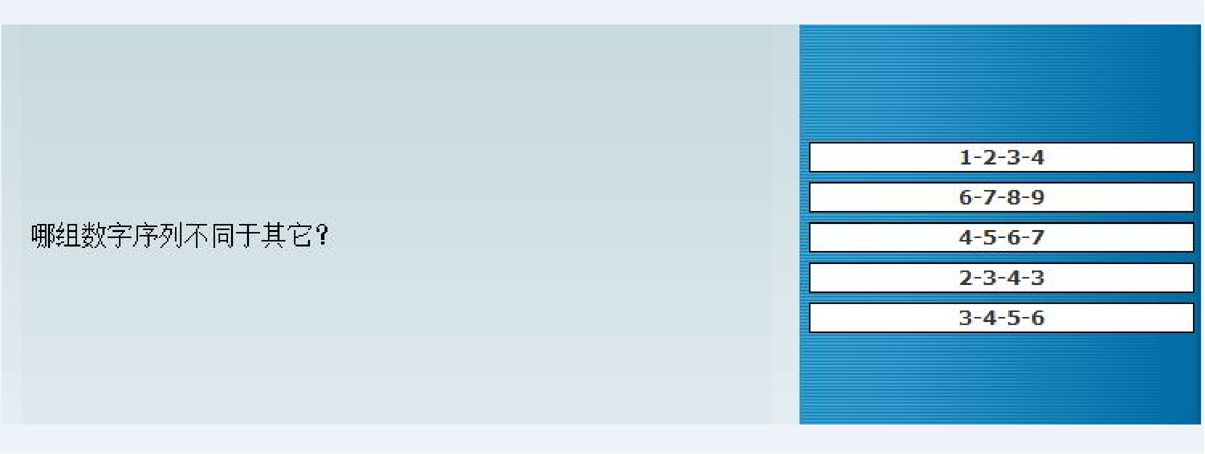


Figure 3.4. Digit sequence task

4.7.2.2 Matrices.

Matrices was also used as a Gf measure in Study 3. Administration of the test was the same as in Study 2 but it was translated into Chinese. For a full description of the task refer to Study 2.

Figure 3.5 provides an illustration of a Matrices item in Chinese.

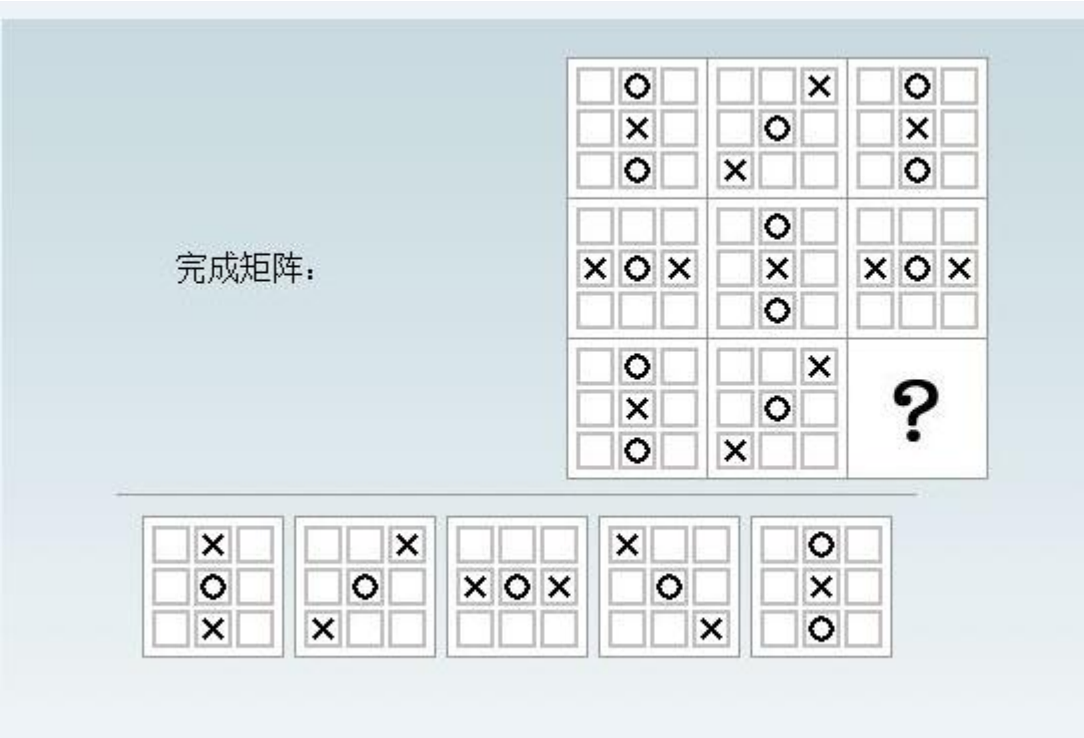


Figure 3.5. Matrices task

4.7.2.3 Reading Comprehension test.

Reading Comprehension was also used as a measure of Gc in Study 3 translated into Chinese. For a full description of this task refer to Study 2. Figure 3.6 provides an illustration of a Reading Comprehension item in Chinese.

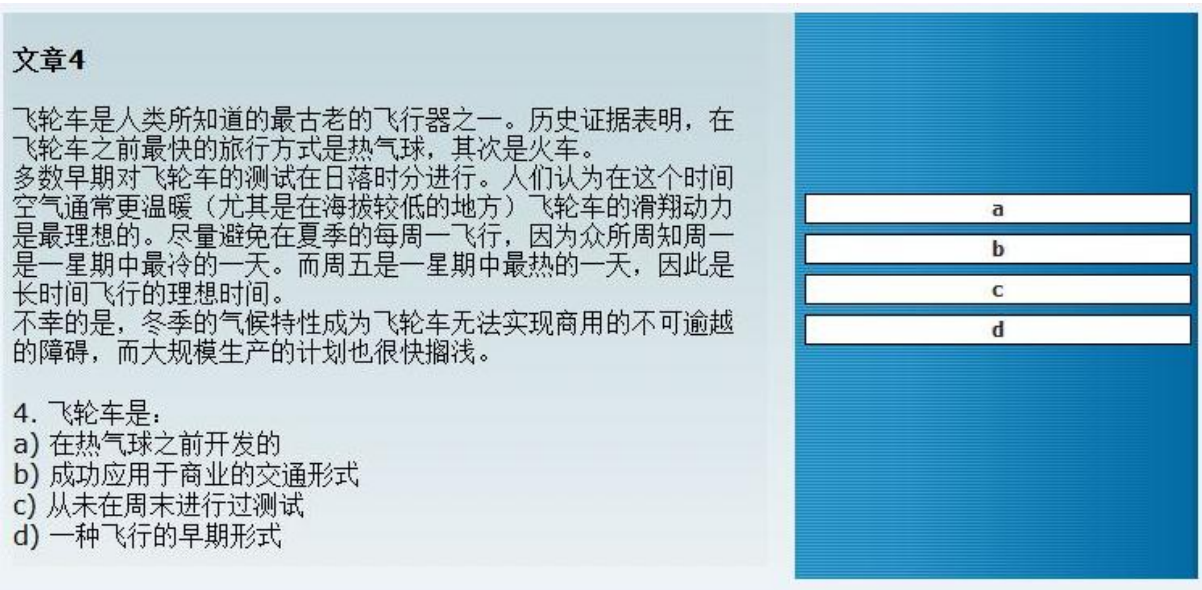


Figure 3.6. Reading Comprehension task

4.7.2.4 Word Reasoning test.

Word Reasoning was used as a measure of Gc. One item was unable to be translated and was not administered. There were 21 items in total and the total accuracy score was therefore 21. In all other aspects administration was the same as the previous studies. Refer to Study 1 for a full description. Figure 3.7 provides an illustration of a Word Reasoning item in Chinese.

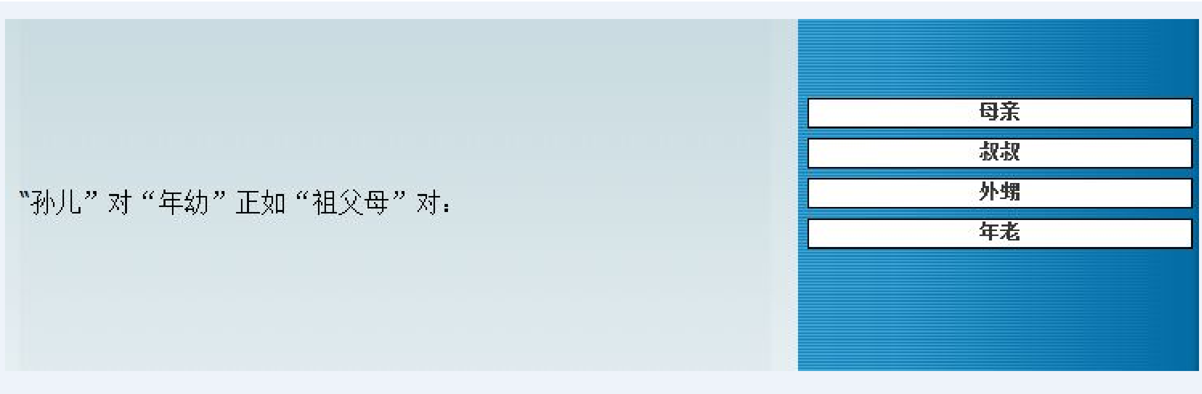


Figure 3.7. Word Reasoning task

4.7.2.5 Number Series test.

In this task, purported to be a measure of Gf, the participants' task is to complete various arithmetic sequences, clicking on their choice from five alternatives. This task is a variant of the Thurstonian series completion test. There were 12 items in total. Responses were scored 0 for incorrect and 1 for correct for each item, the total being the accuracy score out of a possible 12. There was a time limit of 60 seconds for each question. Figure 3.8 provides an illustration of a Number Series item.

Figure 3.8. Number Series task

4.7.2.6 Digit Span test.

Refer to Study 2 for a full description and demonstration of this task measuring Gsm. The items were translated into Chinese for Study 3. No screenshot is provided.

4.7.2.7 Multi-Tasks test.

The administration of the Multi-Tasks test was the same as in the previous studies but the items were in Chinese. See Study 1 for a full description. The sub-tasks in Chinese are illustrated in Figures 3.9 and 3.10.

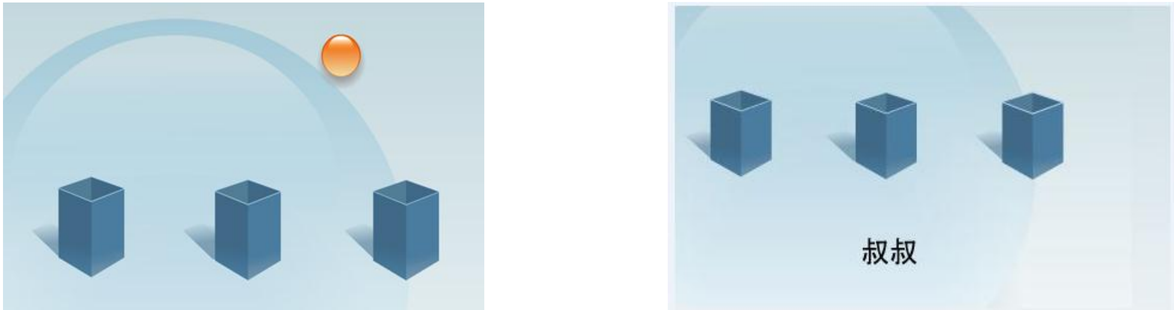


Figure 3.9. Multi-Tasks Placement Keeping component (left) and then Word Recall component (right)

Figure 3.10. Multi-Tasks test answer option whereby the participant must recall the number of balls in each box and then the order of one of the words shown in the display

4.7.2.8 Job Performance measures.

There were three objective job performance measures: a) Absenteeism, as measured by number of days absent, b) Unreliability, as measured by whether participants had been counselled for being unreliable by supervisors (0 = no, 1 = yes) and c) Job Level, which was based on seniority. The participants were all 'operators' within the manufacturing industry. Ratings were based on experience and seniority (1 = least senior/experienced, 5 = most senior/experienced).

Ratings were recorded and collated on Excel spreadsheet by the Human Resources Officer. Ratings of Absenteeism were collected for 58 participants, and of Unreliability and Job Level for 70 participants.

4.7.3 Procedure.

Testing was undertaken via the Internet by invitation from the Human Resources Department. The system was supported by Microsoft.net and Adobe Flash on a Microsoft Sequel Server. The system provided the participants with full instructions, including practice questions (2 questions per sub-task) and feedback if the practice questions were completed incorrectly. Each subject was administered the test battery with the same order of sub-tasks. The test order was as follows: Word Reasoning, Reading Comprehension, Multi-Tasks, Matrices, Number Series, Digit Sequence and Digit Span.

Testing time was approximately 55 minutes. When the tests were completed the system informed the participant that they had completed the assessment. Results were saved to a secure server. The candidates undertook the assessment at the workplace and were supervised by the Human Resources Department.

4.8 Results

In the first section, preliminary procedures will be described. In particular, the outlier checks, reliabilities of the measures, and descriptive statistics will be reported. Following this, independent t-test and analysis of variance results capturing demographic group differences in Multi-Tasks test scores will be described. In the last section, the results of the measurement model and structural equation model tests will be presented.

4.8.1 Preliminary Procedures.

4.8.1.1 Outlier checks.

Total scores were computed for the Multi-Tasks Word Recall, Multi-Tasks Placement Keeping, Matrices, Digit Span, Digit Sequence, Number Series, Word Reasoning, and Reading Comprehension scales. The composites were standardised; cases whose standardised values exceeded the absolute value of 3.29 (per Tabachnick & Fidell, 2007) were considered as outliers. None of the cases had standardised values higher than the absolute value of 3.29; thus, none of the cases were deleted from the data set.

4.8.1.2 Descriptive statistics and reliabilities for the Multi-Tasks subscales.

Word recall. The initial Word Recall alpha was .87, but because one of the items had a negative item-total correlation (i.e., item 2019), it was deleted from the subscale. As shown in Table 3.3, final alpha was .88. The mean score was 4.67 (SD = 4.19).

Placement keeping. The initial alpha for the Placement Keeping subscale was .93. To keep the number of items equal across Word Recall and Placement Keeping measures, item 2019 was dropped. Alpha without this item remained .93. The mean score was 15.10 (SD = 11.75). These descriptive statistics are summarised in Table 3.3.

Table 3.3

Descriptive Statistics and Alpha Coefficients for the Multi-Tasks Subscales

Multi-Tasks Subscale	Item N	Incomplete Data		Complete Data		Alpha
		M	SD	M	SD	
Word recall	15	4.67	4.19	3.60	3.08	.88
Placement keeping	15	15.10	11.75	13.00	10.18	.93

4.8.1.3 Descriptive statistics for the other cognitive ability measures.

Matrices. The initial alpha for the Matrices scale was .47. As three items (i.e., 383, 386, and 387) had item-total correlations between .00 and .10, they were dropped from the scale. Thereafter, alpha increased to .55. The mean score was 3.07 (SD = 1.95).

Digit span. Although the Digit Span scale consists of 14 items, respondents did not answer the last five questions. Alpha for the nine items of the scale was .79. The mean score was 4.50 (SD = 2.72).

Digit sequence. Alpha for this scale was .70. The mean score was 5.74 (SD = 2.74).

Number series. Initial alpha for the Number Series scale was .77. As one item (i.e., 159) had a negative item-total correlation, it was deleted from the scale. Alpha after its deletion was .79. The mean score was 5.41 (SD = 3.14).

Word reasoning. Initial alpha for the Word Reasoning scale was .61. As several items either had negative or very low item-total correlations (i.e., 390, 394, 399, 401, 402, 403, 406, and 1865), they were deleted from the scale. Thereafter, alpha increased to .70. The mean score was 9.59 (SD = 2.45).

Reading comprehension. Initial alpha for the Reading Comprehension scale was .50. Several items that had negative or very low item-total correlations (i.e., 595, 596, 598, 603, 604, 613, 616, and 617) were deleted from the scale. The resultant alpha was .68. The mean score was 4.31 (SD = 2.68). These statistics are given in Table 3.4. Histograms are given in Appendix 3A.

Table 3.4

Descriptive Statistics and Alpha Coefficients for the Other Intelligence Scales (N = 101)

Scale	Item N	Incomplete Data		Complete Data		Alpha
		M	SD	M	SD	
Matrices	9	3.07	1.95	2.88	1.77	.55
Digit span	9	4.50	2.72	4.16	2.72	.79
Digit sequence	12	5.74	2.74	5.95	2.70	.70
Number series	11	5.41	3.14	5.19	3.02	.79
Word reasoning	13	9.59	2.45	9.67	2.18	.70
Reading comprehension	12	4.31	2.68	3.48	1.94	.68

4.8.1.4 Descriptive statistics for the Job Performance measures.

The frequencies and percentages for the job performance variables measured on a nominal or ordinal scale are displayed in Table 3.5. Of the 58 participants for whom information regarding Absenteeism was available, 82.8% had never been absent. Similarly, of the 70 participants who had ratings for the Unreliability item, 87.9% had never been counselled for being unreliable. Lastly, of the 70 participants who provided their job levels, 21.8% were at the third job level, 31.7% were at the fourth job level, and 15.8% were at the fifth job level.

Table 3.5

Job Performance Measures

Variable	Frequency	Percentage
Total absences		
0	48	82.8
1	5	8.6
2	1	1.7
3	2	3.4
4	1	1.7
5	0	.0
6	1	1.7
Unreliability		
0	51	87.9
1	7	12.1
Job Level		
1	0	.0
2	0	.0
3	22	31.4
4	32	45.7
5	16	22.9

4.8.2 Correlations amongst the Variables.

The Pearson correlations between the variables are displayed in Table 3.6. As the Multi-Tasks Word Recall and the Word Reasoning measures were not normally distributed (cf. the histograms in Appendix 3A), they were transformed. The transformed variables were used in the Pearson correlations. The findings reveal that all the intelligence measures were positively correlated with each other. The highest correlations were between Number Series and the other Gf measures and between Reading Comprehension and Multi-Tasks Placement Keeping and Digit Span and Multi-Tasks Word Recall.

Table 3.6

Pearson Correlations amongst the Intelligence Measures (N = 101)

Intelligence Measure	1	2	3	4	5	6	7
1 MT word recall							
2 MT placement keeping	.35***						
3 Matrices	.39***	.43***					
4 Digit span	.54***	.35***	.31***				
5 Digit sequence	.27**	.39***	.37***	.27**			
6 Number series	.31**	.51***	.51***	.46***	.60***		
7 Word reasoning	.31**	.34***	.41***	.32***	.23*	.39***	
8 Reading comprehension	.44***	.57***	.33***	.30***	.31**	.39***	.29**

* $p < .05$ ** $p < .01$ *** $p < .001$

4.8.3 Multi-Tasks Scores across Demographic Groups.

Independent t-test and one-way analysis of variance (ANOVA) procedures were conducted to determine whether scores varied across gender, age, and levels of education. A two-tailed significance level was specified for the independent t-test procedure. A significance level of .05 was specified for the overall model in the ANOVA procedure; when the overall model was statistically significant, post-hoc Tukey procedures were conducted.

4.8.3.1 Gender.

Multi-Tasks Word Recall scores varied significantly across males and females, t (99) = -2.18, p = .036; r^2 = .046. As shown in Table 3.7, males (M = 4.01, SD = 3.47) had significantly lower Word Recall scores than females (M = 6.39, SD = 5.35). Multi-Tasks Placement Keeping scores did not differ significantly across males and females, t (99) = -1.32, p = .192; r^2 = .017.

Table 3.7

Means and Standard Deviations for Multi-Tasks Scores of Males and Females

Gender	N	Word Recall		Placement Keeping	
		M	SD	M	SD
Males	73	4.01	3.47	14.15	10.90
Females	28	6.39	5.35	17.57	13.61

4.8.3.2 Age.

The means and standard deviations for the Multi-Tasks Word Recall and Placement Keeping scores for the various age groups are presented in Table 3.8. The findings reveal that Multi-Tasks Word Recall scores did not vary significantly across age groups, F (6, 94) = .85, p = .539; η^2 = .051. Similarly, Multi-Tasks Placement Keeping scores did not differ across age groups, F (6, 94) = .70, p = .648; η^2 = .043.

Table 3.8

Means and Standard Deviations for Multi-Tasks Scores across Age Groups

Age Group	N	Word Recall		Placement Keeping	
		M	SD	M	SD
18 to 24	16	4.25	4.28	13.19	11.84
25 to 29	30	4.93	3.97	18.17	11.58
30 to 34	20	5.75	4.24	16.10	11.70
35 to 39	20	3.50	4.01	12.35	11.38
40 to 44	9	5.56	4.16	12.67	11.40
45 to 49	4	4.75	6.95	15.75	18.48
50 to 54	2	1.00	0.00	11.50	09.19

4.8.3.3 Levels of education.

The findings in Table 3.9 reveal that Multi-Tasks Word Recall scores varied significantly across levels of education, F (5, 95) = 25.64, p = .001; η^2 = .574. Post-hoc Tukey results indicate that participants with a Bachelor degree (M = 12.44, SD = 3.40) had significantly higher Word Recall scores than those who had a middle school diploma (M = 2.91, SD = 2.43; p = .001), a high school diploma (M = 3.96, SD = 3.21; p < .001), completed technical college (M = 3.63, SD = 2.09; p = .001), and took academic courses (M = 5.00, SD = 4.83; p = .001). Further, participants with a Masters degree (M = 12.40, SD = 2.88) had significantly higher Word Recall scores than those who had a middle school diploma (p = .001), a high school diploma (p = .001), completed technical college (p = .001), and took academic courses (p = .002).

Multi-Tasks Placement Keeping scores also differed significantly across levels of education, F (5, 95) = 10.17, p = .001; η^2 = .349. Post-hoc Tukey results indicate that participants with a Bachelor degree (M = 30.11, SD = 10.45) had significantly higher Word Recall scores than those who had a middle school diploma (M = 11.88, SD = 8.98; p = .001), a high school diploma (M = 13.33, SD = 10.39; p < .001), and completed technical college (M = 10.25, SD = 9.29; p = .001). Further, participants with a Masters degree (M = 33.00, SD = 12.23) had significantly higher Placement Keeping scores than those who had a middle school diploma (p = .001), a high school diploma (p = .001), and completed technical college (p = .001).

Table 3.9

Means and Standard Deviations for Multi-Tasks Scores across Levels of Education

Level of Education	N	Word Recall		Placement Keeping	
		M	SD	M	SD
Middle school diploma	43	2.91	2.43	11.88	8.98
High school diploma	24	3.96	3.21	13.33	10.39
Technical college	4	5.00	4.83	23.50	10.85
Took academic courses	16	3.63	2.09	10.25	9.29
Bachelor degree	9	12.44	3.40	30.11	10.45
Masters degree	5	12.40	2.88	33.00	12.23

4.8.4 Multi-Tasks and Job Performance Measures.

Independent t-test and one-way analysis of variance (ANOVA) procedures were conducted to determine whether scores varied across absenteeism, unreliability, and job levels. A two-tailed significance level was specified for the independent t-test procedure. A significance level of .05 was specified for the overall model in the ANOVA procedure; when the overall model was statistically significant, post-hoc Tukey procedures were conducted.

4.8.4.1 Absenteeism.

As shown in Table 3.10, Multi-Tasks Word Recall scores did not vary across levels of absenteeism, t (56) = .45, p = .653. Similarly, Multi-Tasks Placement Keeping scores did not differ significantly across levels of absenteeism, t (56) = .78, p = .437.

Table 3.10

Means and Standard Deviations for Multi-Tasks Scores across Levels of Absenteeism

Level of Absenteeism	N	Word Recall		Placement Keeping	
		M	SD	M	SD
Zero	48	3.69	3.00	13.47	10.57
One or more	10	3.20	3.55	10.70	8.11

4.8.4.2 Unreliability.

The findings in Table 3.11 reveal that Multi-Tasks Word Recall scores did not vary across levels of unreliability, t (56) = -.89, p = .380. Similarly, Multi-Tasks Placement Keeping scores did not differ significantly across levels of unreliability, t (56) = -1.36, p = .181.

Table 3.11

Means and Standard Deviations for Multi-Tasks Scores across Levels of Unreliability

Level of Unreliability	N	Word Recall		Placement Keeping	
		M	SD	M	SD
Zero	51	3.47	3.00	12.33	9.62
One	7	4.57	3.69	17.86	13.46

4.8.4.3 Job Level.

The findings in Table 3.12 reveal that Multi-Tasks Word Recall scores varied significantly across job levels, F (2, 67) = 20.97, p = .001. Post-hoc Tukey results indicate that participants at the fifth job level (M = 10.63, SD = 5.37) had significantly higher Word Recall scores than those in the third (M = 3.77, SD = 3.12; p = .001), and fourth job levels (M = 3.66, SD = 3.16; p = .001).

Multi-Tasks Placement Keeping scores also varied significantly across job levels, F (2, 67) = 18.63, p = .001. Post-hoc Tukey results indicate that participants at the fifth job level (M = 30.00, SD = 10.70) had significantly higher Placement Keeping scores than those in the third (M = 12.00, SD = 10.70; p = .001), and fourth job levels (M = 12.50, SD = 9.57; p = .001).

Table 3.12

Means and Standard Deviations for Multi-Tasks Scores across Job Levels

Job Level	N	Word Recall		Placement Keeping	
		M	SD	M	SD
Third	22	3.77	3.12	12.00	10.70
Fourth	32	3.66	3.16	12.50	9.57
Fifth	16	10.63	5.37	30.00	10.70

4.8.5 Results of Measurement Model Tests.

4.8.5.1 Creation of parcels.

Parcels (i.e., small item groups) were created for two reasons. Firstly, the tests were scored using a nominal scale (i.e., incorrect vs. correct) and because a requirement of structural equation modeling is that the data be measured on either an interval or ratio scale, items had to be summed into parcels. Secondly, as Little et al., (2002) identified, models with single-item indicators are less parsimonious and often increase sampling error. Therefore, parcels were created.

Multi-Tasks subscales. The Multi-Tasks items were arranged in order of difficulty. Parcels were thus created based on number order. Four parcels consisting of four items were created; the first parcel had the least difficult items while the fourth parcel had the most difficult items (refer to Appendix 3B for the parcel compositions).

Intelligence measures. As items for these tests were not arranged in order of difficulty, the item-to-construct balance method, recommended by Little et al., (2002) was used to create parcels. Specifically, the corrected item-total correlations (from the reliability analyses) were used to generate the parcels. These correlations were sorted from highest to lowest. Items with the highest item-total correlation anchored each of the parcels. The items with the next highest item-total correlations were added to the anchors in the reverse order. The item with the highest correlation among the anchor items was matched with the lowest loading item from the second selection. This basic procedure where items with lower correlations were matched with items with higher correlations was repeated until all items were categorised into parcels. Refer to Appendix 3B for the parcel compositions.

4.8.5.2 Assessment of multivariate normality.

Multivariate normality is an important assumption of structural equation modeling; thus normality of the parcels was assessed. According to Kline (2005), multivariate normality can be assumed when the univariate distributions are normal and the distribution of any pair of variables is bivariate normal. Kline (2005) further proposed that because it is impractical to examine all joint distributions, examining univariate distributions will usually allow one to detect instances of non-normality. Thus, the skewness and kurtosis of the parcels were assessed.

The skew indices of some of the parcels were above the acceptable criterion of three (Kline, 2005). Accordingly, these variables were transformed to correct for skewness (Howell, 1992). The variables that were negatively skewed were normalised using a power transformation; the variables that were positively skewed were normalised using a square root transformation. These transformed variables were used in subsequent analyses.

4.8.5.3 Assessment of model fit.

Model fit was assessed by interpreting several fit indices including the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), the Standardised root mean residual (SRMR), and the likelihood ratio χ^2 test. A model is deemed as fitting the data well when the CFI value is above .95 (Hu & Bentler, 1999). Browne and Cudeck (1993) suggest that a model with an RMSEA value less than .05 has good fit, one with a value than .08 has reasonable fit, and a model with an RMSEA less than .10 has poor fit. Kline (2005) revealed that a model with an SRMR value of less than .10 has good fit. A small χ^2 value relative to the degrees of freedom (i.e., values lower than 3) indicates good model fit (Hu & Bentler, 1999).

Nested models were compared via the change in chi-square. In addition, two information indices were reported: the Akaike Information Criterion (AIC) and the Expected Cross-Validation Index (ECVI). Information indices are relative; models indices are compared to each other and the lower the value, the better the model fit (Byrne, 2001).
In addition to evaluating the model as a whole, the significance of the individual parameters was also assessed (Byrne, 2001). Parameters were evaluated at the .05 level. The direction of the standardised path coefficients was checked to see if it was consistent

4.8.5.4 Confirmatory factor analysis for the Multi-Tasks measurement model.

The proposed two-factor model using the transformed Multi-task parcels is depicted in Figure 3.11. This two-factor model was compared to a single-factor model. The fit indices for both models are summarised in Table 3.13 and reveal that the proposed two-factor model fit the data better than the single-factor model. Firstly, the fit indices of the two-factor model met all the criteria for good fit: the Normed chi-square was below two, the CFI was above .95, the RMSEA was below .06, and the SRMR was low at .05. Secondly, the change in chi-square between the two-factor and single-factor model was statistically significant, $\Delta\chi^2$ (1) = 150.03, $p < .001$. Lastly, the AIC and ECVI values of the proposed two-factor model were much lower than the values of the single-factor model. The maximum likelihood estimates for the two-factor model are displayed in Appendix 3C. This two-factor model was thus carried through to the next stage of the analysis.

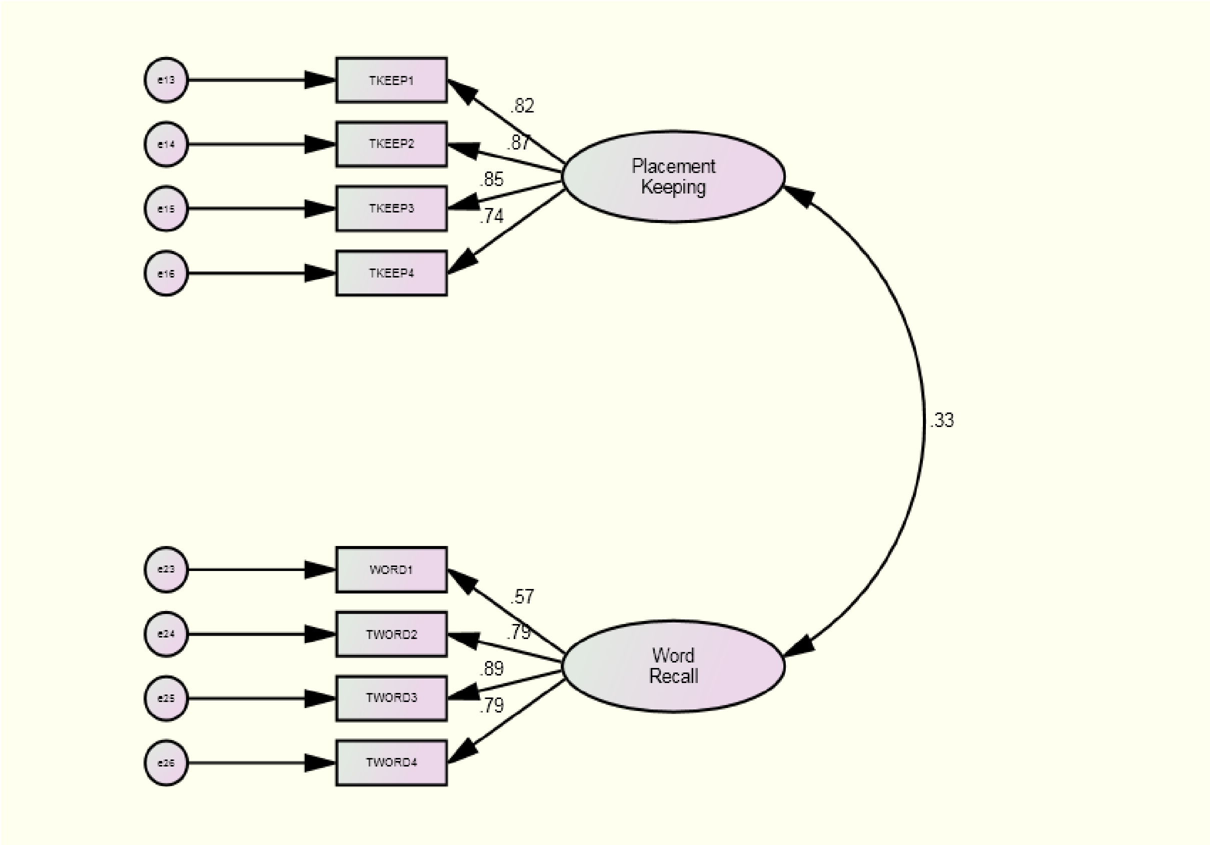


Figure 3.11. Results for the proposed two-factor model (with standardised coefficients).

Table 3.13

Chi-square Results and Fit Indices for the Multi-Tasks Measurement Models

Index	Single Factor	Two-Factor
Chi-square	173.65	23.62
Degrees of freedom	20.00	19.00
Sig.	.00	.21
Normed chi-square (chi-square/df)	8.68	1.24
Comparative fit index (CFI)	.63	.99
Root mean squared error of approximation (RMSEA)	.28	.05
Lower bound 90 % CI	.24	.00
Upper bound 90 % CI	.32	.11
Standardised root mean residual (SRMR)	.17	.05
Akaike information criterion (AIC)	205.65	57.62
Expected cross-validation index (ECVI)	2.06	.58

Note. Critical χ^2 (1) = 10.83, p < .001.

4.8.5.5 Results for the confirmatory factor analysis of the full measurement model.

Prior to testing the proposed second-order measurement model, a first-order measurement model was tested. Only one parcel, the first Word Reasoning parcel, was deleted from the model (as it was highly correlated with the Reading Comprehension construct). The revised first-order measurement model was then used in the proposed full second-order two-factor measurement model depicted in Figure 3.12.

The proposed second-order two-factor model was compared to two other models: a second-order single-factor model (depicted in Figure 3.13) and a second-order three-factor model. The second-order three-factor model yielded a non-positive definite matrix; thus, its solution was not admissible. The proposed second-order two-factor model did not fit the data better than the more parsimonious second-order single-factor model: although both models met all the criteria for good model fit, the change in chi-square between the proposed model and the more parsimonious model was not statistically significant, $\Delta\chi^2$ (1) = .07, NS. The path coefficients from second-order to first-order constructs for the parsimonious and proposed models are presented in Appendix 3D. As the two-factor Multi-Tasks model was used in the first and second studies, and fits with the CHC Theory, the two-factor Multi-Tasks model was carried through in subsequent tests.

Table 3.14

Chi-square Results and Fit Indices for the Full Measurement Models

Index	Single Factor	Two-Factor
Chi-square	415.73	415.66
Degrees of freedom	342.00	341.00
Sig.	.00	.00
Normed chi-square (chi-square/df)	1.22	1.22
Comparative fit index (CFI)	.93	.93
Root mean squared error of approximation (RMSEA)	.05	.05
Lower bound 90 % CI	.03	.03
Upper bound 90 % CI	.06	.06
Standardised root mean residual (SRMR)	.08	.08
Akaike information criterion (AIC)	543.73	545.66
Expected cross-validation index (ECVI)	5.44	5.46

Note. Critical χ^2 (1) = 3.84, $p < .05$.

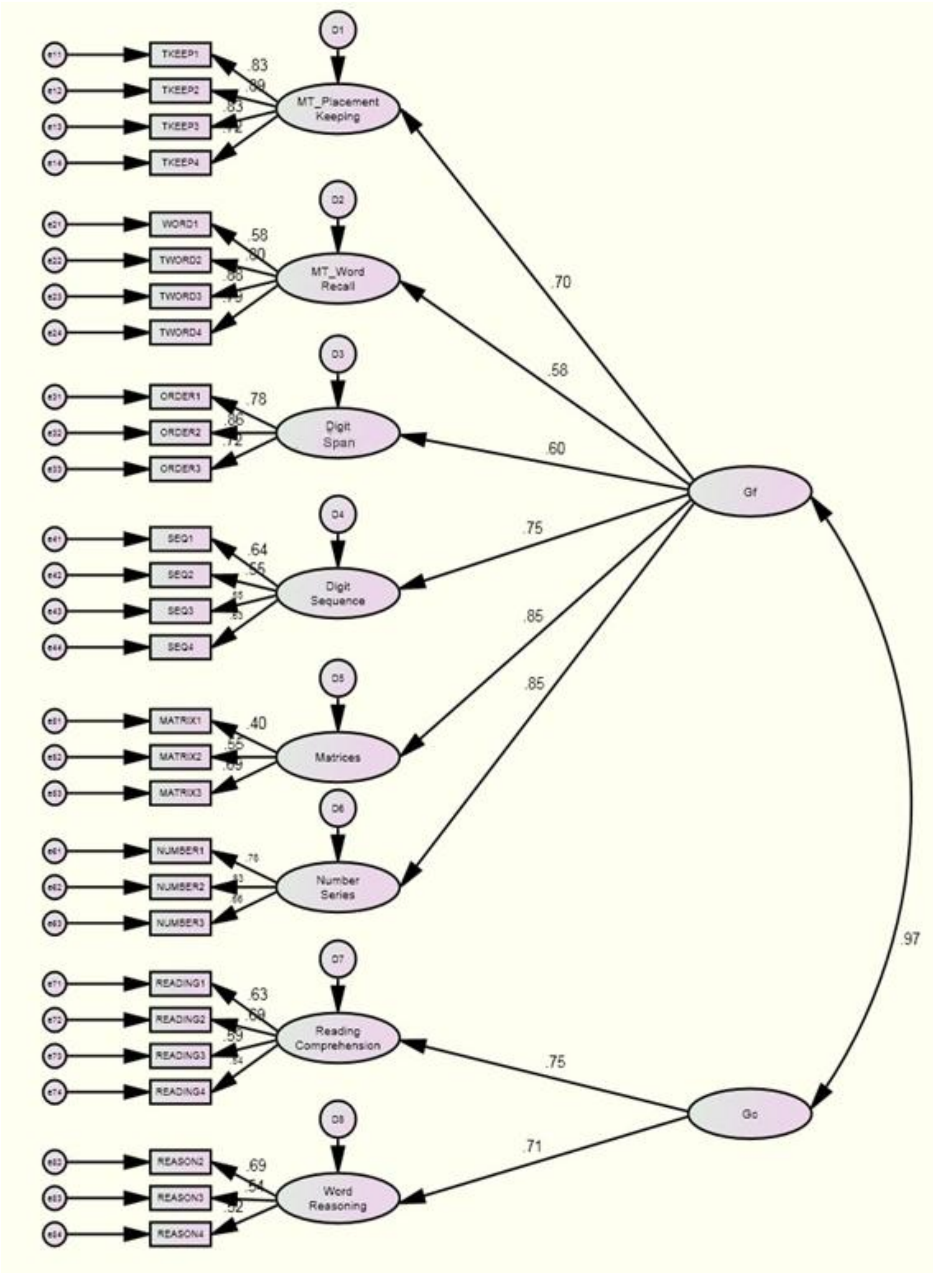


Figure 3.12. Results for the proposed second-order two-factor full measurement model (with standardised coefficients).

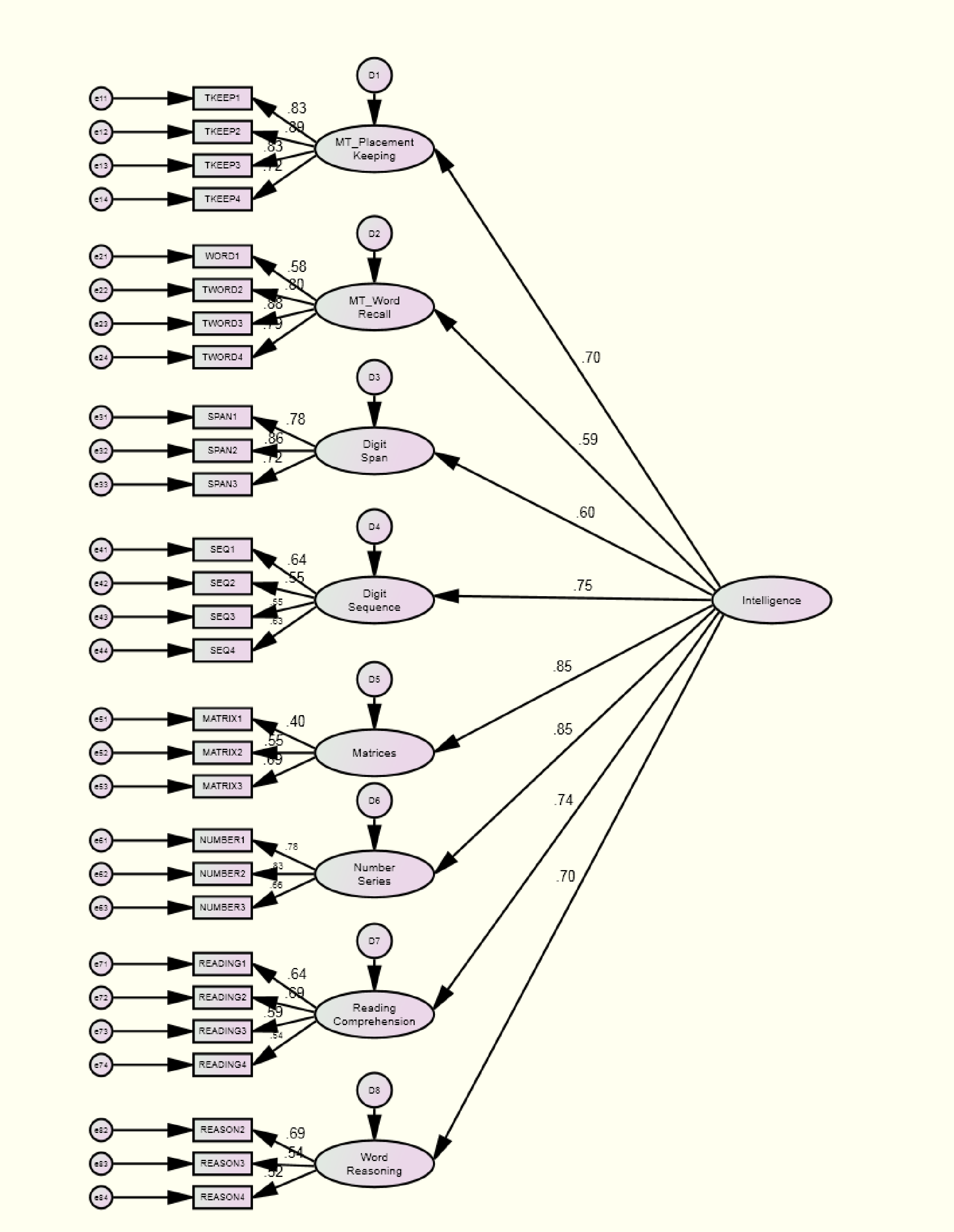


Figure 3.13. Results for the alternative second-order single-factor full measurement model (with standardised coefficients).

4.8.6 Predictive Validity of the Intelligence Measures.

To determine the predictive validity of the Multi-Tasks subscales and the other intelligence measures, Kendall Tau correlations, linear regression, and one-way ANOVA procedures were conducted. Because the job performance measures were measured either on a nominal or ordinal scale, Kendall Tau correlations were first conducted between the intelligence measures and the three job performance measures. If the Kendall Tau correlation between the intelligence measures and the job performance measure was statistically significant, further multivariate linear or logistic regression procedures were conducted. Finally, the job performance measures were combined to yield one job performance measure. The total number of absences and unreliability were deducted from the job level score to yield this job performance measure. The sample was then split into three groups via a tertile split; one-way ANOVA procedures were then conducted with the various intelligence measures and this categorical job performance variable.

4.8.6.1 Correlations between intelligence and Job Performance.

The Kendall Tau correlations presented in Table 3.15 reveal that none of the intelligence measures were significantly correlated with total number of absences and unreliability. The Multi-Tasks Word Recall subscale was positively associated with job level ($\tau = .38$, $p = .001$). The Multi-Tasks Placement Keeping subscale was also positively correlated with job level ($\tau = .49$, $p = .001$). Further, the Matrices ($\tau = .26$, $p = .029$), Digit Sequence ($\tau = .34$, $p = .004$), Number Series ($\tau = .34$, $p = .004$), and Reading Comprehension ($\tau = .56$, $p = .001$) scales were also positively associated with job level.

Table 3.15

Kendall Tau Correlations between the Intelligence Measures and Job Performance (N = 58)

Intelligence Measure	Absences	Unreliability	Job Level[1]
MT word recall	-.06	.12	.28**
MT placement keeping	-.04	.14	.37***
Matrices	.02	.12	.21*
Digit order	-.09	.17	.17
Digit sequence	.13	.05	.29**
Number series	-.01	-.08	.26**
Word reasoning	-.03	.16	.12
Reading comprehension	.01	-.08	.38***

[1] N = 70.

* $p < .05$ ** $p < .01$ *** $p < .001$

4.8.6.2 Predictors of Job Level.

As several of the intelligence measures were positively associated with Job Level, a multiple linear regression procedure was conducted. The findings in Table 3.16 reveal that, after controlling for the effects of the other intelligence measures, only Reading Comprehension significantly predicted job level ($\beta = .36$, $p = .012$). Reading Comprehension accounted for 12.67% of the variance of job level. The proportion of variance accounted for by this model was 39%.

A second regression procedure was conducted where job level was regressed on a single intelligence composite measure. The intelligence composite was the sum of the weighted intelligence tests (where weights from an exploratory factor analysis were used). The findings of this procedure, summarized in Table 3.17, indicate that the intelligence composite significantly predicted job level. The proportion of variance accounted for by this model was 28%.

Table 3.16

Multiple Linear Regression Results for Intelligence and Job Level (N = 70)

Intelligence Measure	β	F	Sig.	r^2	TOL
MT word recall	.18	1.76	.190	.03	.55
MT placement keeping	.23	2.80	.099	.05	.51
Matrices	-.07	.28	.598	.00	.57
Digit order	-.03	.04	.847	.00	.53
Digit sequence	.11	.82	.369	.01	.64
Number series	.06	.14	.707	.00	.43
Word reasoning	-.17	1.67	.202	.03	.60
Reading comprehension	.36	6.63	.012	.13	.52

Note. TOL = Tolerance. Overall model F (8, 61) = 4.96, p = .001. Overall model R^2 = .39.

Table 3.17

Multiple Linear Regression Results for Weighted Intelligence Measure and Job Level (N = 70)

Variable	β	F	Sig.	r^2
Intelligence composite	.53	26.68	.000	.28

4.9 Discussion

The main aim of the current study was to explore the cross-cultural validity of the Multi-Tasks test and to compare Multi-Tasks performance with job performance in another culture, namely China. Specific aims were to: (1) determine whether the Multi-Tasks was affected by culture by examining the performance of Chinese workers within the Chinese workforce, (2) determine whether the Multi-Tasks test predicted job performance within a Chinese context, (3) determine if Multi-Tasks replicates a two factor theoretical model as proposed in Study 2 and examine its relationship with other intelligence measures, and (4) explore demographic differences, specifically whether performance varies by age, gender, job level and level of education.

As expected, the underlying structure of the Multi-tasks test was found to consist of two factors, Word Recall and Placement Keeping. In this study all the intelligence measures, including the Multi-Tasks sub-tasks were positively correlated with each other. As expected, Word Recall was highly correlated with the Gsm measure (Digit Span). These two tasks share similarities in the types of mental procedures used to solve them. They both require memory of lists presented sequentially. Further, Gsm is known to be a basic component of working memory ability and Gf relies on strong working memory skills, thus both these factors are related to the ability to hold in mind and manipulate information (McGrew, 2004; 2009; Stankov, 1988), and both have been shown to be important in competing tasks performance (Conway et al., 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, & Conway, 2005).

The Multi-tasks subtask Placement Keeping was highly correlated with the Number Series and Matrices tests. The Placement Keeping component of Multi-tasks involves elements of deductive problem-solving similar to that of both the Number Series and Matrices tests. The Placement Keeping subtask requires respondents to decide how many balls remain in each of three boxes following a sequence of moves. The Number Series task requires respondents to determine the next series in a sequence, and the Matrices task operates similarly albeit with more complexity. The Number Series task was included in this study as an additional numerical test of Gf as the previous studies showed the reliability coefficient of Digit Sequence to be relatively low (.59 and .46 respectively) and lower than that of the other measures. In hindsight, given the higher reliability of Digit Sequence in the current study (.70) Number Series was not required. It is however interesting to examine how it operated in relation to Multi-Tasks.

Candidates in this study were all tightly supervised whilst undertaking the assessment. This helped to ensure that the results were valid and there was no opportunity for candidates to take notes to maximise their performance. This was particularly important as the negatively skewed variables in Study 2 may have been due to inconsistent test invigilation practices.

This appeared to be an issue isolated to Study 2, as the variables were normally distributed in both Study 1 (see Appendix 1A) and the current study. (see Appendix 3A).As a result of high correlations between the measures of intelligence two models were tested and fitted the data well. The first was a second-order two factor model with Multi-Tasks, the Gf markers and one Gsm marker loading onto Gf factor and the two Gc markers loading onto to Gc factor. This replicated the model in Study 1. The second model, a second-order, one factor model was also acceptable with all intelligence measures loading onto a single intelligence factor (g). It is not surprising that the one factor model fitted the data well considering the strong correlation between all the intelligence measures. This also supports previous research by Kyllonen (1993) and Stauffer, Ree and Carretta (1996) who found strong correlations between each of several psychometric measures of cognitive ability and an overall factor (g). In the Stauffer et al. (1996) study, the correlations between each factor and g ranged from .82 to .98. These were stronger than the correlations between the individual measures for all 35 tasks, other than between two measures of arithmetic and two measures of verbal ability.

A separate Gsm factor could not be identified in the study (as was found in Study 2), with a second order three factor model not fitting the data well. This result is most likely due to the high correlation between the Gsm marker Digit Span and Multi-tasks.

The strong relationship between Multi-Tasks Placement Keeping and the purported Gc marker Reading Comprehension was somewhat surprising. This finding may be related to the demands both tasks place on working memory. The Reading Comprehension task places significant demands on working memory store because the respondent needs to remember chunks of information in order to make accurate conclusions.

Group differences revealed an effect for Education level on Multi-Tasks performance, whereby more highly educated participants had higher scores on both sub-tasks. A general pattern appears to be emerging whereby more educated, older and more experienced respondents perform better on the tasks. While it might be expected that seniority would be confounded with age, this was found to be non-significant ($r = -.043$, $p > .05$). The participants appear to have been hired and promoted on factors mostly unrelated to age. There was also an effect for gender, with males scoring lower than females on Word Recall. However, this finding may not be practically significant due to the relative greater portion of males who undertook the assessment.

One of the most encouraging findings was discovered when the analysis of variance procedure revealed that Multi-tasks results varied significantly across job level. More senior respondents scored better on Multi-Tasks. Job level is a key benchmark of success at work and this finding is very promising. The strong correlations between Multi-Tasks and Job Level supports research indicating that competing task paradigms are particularly useful for predicting performance in complex and managerial jobs (Ben-Shakhar & Sheffer, 2001; North & Gopher, 1976; Sauer et al., 1999; Stankov et al., 1989). The strong relationship between Multi-Tasks and overall intelligence also supports the idea that Multi-Tasks would be related to Job Level, as general intelligence has been shown to be strongly associated with occupational level, with correlations as high as .95 (Gottfriedson et al., 1997; Jensen, 1998; Judge, Higgins, Thoresen & Barrick, 1999).

Deeper analysis did not uncover more detail about the predictive relationship between Multi-tasks and Job Level. A multiple linear regression analysis showed Reading Comprehension was the only significant predictor of job level when controlling for the effects of the other intelligence measures. The positive result for the Reading Comprehension tests supports the notion that language reasoning skills are important for success at work and clearly important for more senior roles.

Counterproductive measures of job performance were examined to determine if Multi-tasks could be useful as a risk management tool to identify unreliable employees. It was found that none of the problem solving measures predicted negative outcomes. These findings may reflect that counterproductive measures of job performance are not the preferred measures to use when examining the relationship between cognition and job performance. This was unavoidable; however, as these were the measures that were used to judge work performance within this occupational setting. It is also reasonable to assume that there would be a relationship between job performance as assessed with more traditional measures and counterproductive measures (i.e., high performing employees are likely to have low levels of absenteeism and unreliability). Further, there is emerging evidence of a possible relationship between cognition and absenteeism. Henderson, Richards, Stansfeld and Hotopf (2012) found a relationship between intellectual ability in childhood and later work absenteeism in a group of 32,660 adults in the United Kingdom from three birth cohorts: 1946, 1958 and 1970 that was not wholly accounted for by ill-health. From these three cohorts, between 31 and 47 percent of those with long periods of absenteeism were in the lowest quartile in terms of intelligence (Henderson et al., 2012). Further research could look at the relationship of current intellectual ability with absenteeism and the efficacy of personality assessment in predicting counterproductive behaviours such as absenteeism or other work adjustment problems.

Limitations of the current study included characteristics of the sample and the job performance measures used. The sample size was relatively small (particularly in comparison to Study 1 and Study 2), there was an over-representation of males and individuals under the age of 35 years and no participants at job levels one or two. A greater spread of ages and job levels may have better revealed the way Multi-Tasks operates within the Chinese workforce and it would have been preferable to have equal numbers of males and females to minimise any gender effects. As previously discussed, traditional measures of job performance may have been preferable to counterproductive measures for the purposes of this study. A limitation of involving organisations in research is that a researcher has limited influence as to the characteristics of the sample and the job performance measures available. This is particularly the case in cross-cultural research as there is limited opportunity to be directly involved in selecting the sample. It was thought that the advantages of undertaking this research within organisations, rather than in a better controlled environment, outweighed the disadvantages as it provided an opportunity for the researchers to investigate how Multi-Tasks operates within a naturalistic work setting. Further, characteristics of the sample were reflective of the organisation: more males and younger individuals characterised the organisation involved in this study. While a larger sample size would have been preferable, the outcomes of the current study were consistent with Study 1 and reasonably consistent with Study 2, therefore there is no reason to believe that the results were adversely affected by a low sample size.

The present study was the first of its kind to investigate the cross-cultural validity of the Multi-Tasks test. The results showed the test to be a valid and highly reliable measure within the Chinese culture, with outcomes that were comparable to those of the Australian participants in the previous studies. Chinese employers may benefit from administering a battery of problem solving tests and especially Multi-Tasks and a Reading Comprehension test because they show promise in identifying those individuals with potential for senior roles.

CHAPTER 5

GENERAL DISCUSSION

COMPETING TASKS AS MEASURES OF INTELLIGENCE

AND PREDICTORS OF JOB PERFORMANCE

General Discussion

Studies 1, 2 and 3 suggest that the Multi-Tasks test is a reliable and valid measure of performance at work, and may be especially useful for predicting an individual's potential for senior level roles. It is also apparent that verbal reasoning measures are important for success at work. A person's capacity to understand language and express themselves effectively is clearly relevant to many jobs, particularly those at senior levels and cross-culturally. Specific study outcomes will be discussed, commencing with group differences.

5.1 Gender

There was an effect for gender on performance on one component of the Multi-Tasks test in two of the studies, Study 1 and Study 3. Study 1 (Part B) found that males outperformed females on the Placement Keeping sub-task while there was no gender difference on the Word Recall sub-task. In Study 3, females outperformed males on the Word Recall sub-task and there was no difference on Placement Keeping. There was no effect for gender on either sub-task in Study 2. Thus, there appears to be no consistent pattern with regard to gender differences. Similarly, earlier research on competing tasks paradigms has found no clear gender differences (e.g., Hofheinz et al., 2010), although many researchers in this area have not explored gender differences. Further research would be required to investigate whether there are clear gender differences, or to confirm the lack of an effect for gender. This research would prove useful in determining whether there should be separate norms for males and females on the Multi-Tasks test.

5.2 Education

An effect for education was found in the pilot study (Study 1 Part A), Study 2 and Study 3. The pilot study found an effect for level of education on the Placement Keeping sub-task and study 2 found an effect for level of education on Multi-Tasks performance in the initial ANOVA, however post-hoc tests were non-significant. In Study 3 more highly educated participants had higher scores on both sub-tasks of the Multi-Tasks test. Initial findings therefore suggest that more educated respondents may perform better on the Multi-Tasks test. Previous research has shown that competing task paradigms have a stronger relationship with general intelligence than single tasks (e.g., Stankov, 1988) and that competing tasks are better predictors of management potential than the same tasks presented singly (e.g., Stankov, et al., 1989). Further, in the current group of studies an effect for organisational level was found in Study 1 and Study 3, whereby more senior employees or applicants for more senior roles performed better on the Multi-Tasks test. It is therefore unsurprising that initial findings show a positive relationship between education and test performance, as managers and more intelligent individuals are likely to have higher levels of education.

5.3 Age

An effect for age was only found in Study 1 (Part B), whereby older participants had higher scores on both Multi-Tasks sub-tasks. No effects for age were found in the pilot study or Studies 2 or 3; however this may be due to the over-representation of younger people in studies 2 and 3. Research shows that performance on Gf tasks tend to decline with age (Horn, 1980; Horn & Stankov, 1982; Stankov, 1986), however Stankov's (1986) study demonstrated that competing tasks paradigms are more resistant to the effects of ageing than single tasks. It should be noted that the age at which Multi-Tasks test performance was best was between 40 to 44 years. . This is interesting, because the age at which Multi-Tasks test performance was optimal is the age at which individuals are likely to adopt high level managerial roles.

Effect sizes for each of the demographic variables across all three studies were also examined. These are outlined in the following section.

5.4 Demographic effects between the studies

Table 4.1 shows the results of the analysis of the average effect sizes for each demographic variable across all the studies. As shown in Table 4.1, Multi-Tasks scores did not vary across males and females. The average effect size for the Word Recall subscale was 1.5% and the average effect size for the Placement Keeping subscale was 2.5%. The proportion of variance explained by levels of education was higher, 19.9% for Word Recall and 12% for Placement Keeping. Note, however, that this was primarily due to differences within the Chinese sample (i.e., Study 3). Multi-Tasks scores also did not vary greatly across age groups. The average effect size for the Word Recall task was 2.9% and the average effect size for the Placement Keeping task was 3.1%. Cohen (1988) advises that for r^2 an effect size of .01 shows a small effect, .09 shows a medium effect and .25 a large effect. For η^2 the corresponding figures are .01 (small), .06 (medium) and .14 (large). Level of education was the only demographic variable that had a large effect size, and this provides further support for the suggestion that Multi-Tasks is a valuable predictor of job performance for more complex and managerial roles.

Table 4.1

Effect Sizes for Multi-Tasks Scores across Gender, Levels of Education, and Age Groups

Variable	Study 1		Study 2		Study 3		All Studies	
	WR	PK	WR	PK	WR	PK	WR	PK
Gender[1]	.000	.057	.000	.002	.046	.017	.015	.025
Level of education[2]	.020	.001	.004	.011	.574	.349	.199	.120
Age group[2]	.036	.049	.000	.000	.051	.043	.029	.031

Note. WR = Word Recall. PK = Placement Keeping.

[1] Effect size index was r^2.

[2] Effect size index was η^2.

Statistical properties of the measures and the influence of this on study outcomes will now be considered.

5.5 Correlations between tests and Confirmatory Factor Analysis

Patterns of correlations between tests are useful in providing indications as to what outcomes might be expected from confirmatory factor analysis. Studies 2 and 3 had the greatest number of problem solving measures and provide the best illustration of the utility of correlation matrices in helping guide and explain research outcomes. In study 2, strong correlations between Multi-Tasks and Digit Span indicated that those two tests would load onto a single factor. Strong correlations between the Gc measures (Word Reasoning and Reading Comprehension) and the Gf measures (Matrices and Digit Sequence) suggested that it was also possible, while not expected, that they might load onto a single factor. While it was somewhat surprising that the Gc measures could not be easily distinguished from the Gf measures, this finding can be explained by the Gc tasks, particularly Reading Comprehension placing demands on working memory by requiring details of a passage to be remembered when answering questions. Research shows a strong relationship between working memory load and Gf (Bühner & Mürling, 2005; König, et al., 2005; Stankov, 1988), and in one study a relationship between Gc and working memory (Stankov & Myors, 1990). The revised measurement model in Study 2 is a legitimate outcome when these factors are considered.

In Study 3, all of the problem solving tests were significantly correlated. It was therefore reasonable to expect that a single factor model might have fitted the data well and did so. In Study 1 and in one of the models for Study 3, the Gc measures were defined as a separate factor as expected, however in Study 3 the Gsm measure and the Gf measures loaded onto the one factor. This was due to the strong correlations between the variables. Further, it is the author's experience that differentiation between measures of Gsm and Gf are at times difficult to obtain. This is likely due to the fact that working memory is closely related to Gsm and also an important aspect of Gf.

The current series of studies did not confirm the existence of a time-sharing factor. Importantly however, the protocol used in this series of studies makes it more difficult to identify a time-sharing factor. Unlike previous protocols (e.g., Stankov, 1988), Multi-Tasks was not administered in its individual components due to restrictions on testing time. In previous research, two tasks were administered independently and then paired. This allows for additional comparisons between single and dual task performance to be made. While the author was interested in both the tests and the constructs underpinning the tests, the constructs tend to become less relevant if there is no construct associated with competing tasks, such as the so-called timesharing factor.

In summary, high correlations between certain tests as well as the reasoning component of tests can make it difficult for them to emerge as separate factors. It should also be noted that the coefficients were not corrected for attenuation, and therefore the relationships among the variables are higher than reported.

5.6 Key statistical strengths of Multi-tasks

In all studies, the reliability coefficients for the Multi-Tasks sub-tests were high (ranging from .79 to .93) and higher than those of the other cognitive ability measures. Reliability is as an important criterion for determining the quality of a problem solving test (Tabachnick & Fidell, 2007). Multi-Tasks had a consistently strong relationship with Gf, particularly in Studies 1 and 3 and an association with Gsm in Study 2. Gsm is known to be a component of Working Memory (McGrew, 2004; 2009) which has been shown to be an important component in performance on Gf tasks. These outcomes support previous research demonstrating the relationship between competing tasks measures, Gf and WM (Bühner & Mürling, 2005; König, et al., 2005; Stankov, 1988).

One of the main aims of the studies was to investigate the relationship between Multi-Tasks and job performance. The outcomes of this analysis will be discussed in detail.

5.7 Job performance outcomes

In Study 1 (Part B) Multi-Tasks correlated strongly with two supervisor ratings, Complex Problems and Promotional Potential, which themselves were highly correlated suggesting supervisors rated respondents in a similar way across these dimensions. The Gc measure (Word Reasoning) and Gf measure (Digit Sequence) were also associated with at least two of the three supervisor ratings, but Multi-Tasks was the strongest predictor.

In terms of objective measures of job performance, correlation analyses in Study 2 revealed that Multi-Tasks was significantly related to Years of Service, as were two Gf measures. Logistic regression analyses showed the Word Recall Multi-Tasks sub-task to be predictive of Voluntary Transfer and both sub-tasks to be predictive of Job Level when controlling for the effects of the other measures. Matrices (a Gf measure) was also predictive of Job Level. When the three objective measures were summed to create an overall measure of job performance, Multi-Tasks, Reading Comprehension (Gc measure) and Digit Span

(Gsm measure) showed significant results with better performance on these tasks being associated with better job performance as indicated by being with the company for longer, more likely to qualify for voluntary transfer and higher job level.

Study 3 revealed that Multi-Tasks, the Gf measures and Reading Comprehension (Gc measure) were significantly related to the Job Level performance indicator, while none of the cognitive ability measures were associated with the other indicators, Absenteeism or Unreliability. In the same study, regression analyses showed Reading Comprehension to be the only significant predictor of Job Level when controlling for the effects of the other intelligence measures. This finding highlights the importance of language reasoning skills within the work context.

Multi-Tasks was designed to be applicable and accessible to a wide range of job types and industries without the need for sophisticated equipment or tailoring to particular job roles. The results across the three studies provide support for this: in three very different work contexts, including a Chinese work context, Multi-Tasks was found to be a consistently strong predictor of at least some indicators of job performance. The Reading Comprehension task is also a strong predictor of job performance.

The relationship between Multi-Tasks and Job Level supports research indicating that competing task paradigms are particularly useful for predicting performance in complex and managerial jobs (Ben-Shakhar & Sheffer, 2001; North & Gopher, 1976; Sauer et al., 1999; Stankov et al., 1989).

5.7.1 Quality of job performance data.

Job performance data was not available for all participants and there were inconsistencies around the way job performance ratings were collected. While providing clear ratings guidelines and consistent standardised measures between studies would have been favourable, this was not possible as the organisations in Study 2 and 3 had existing performance frameworks and were not in a position to re-rate participants on an additional performance framework. In Study 1 (Part B), ratings guidelines were provided by the author and it was in this study that Multi-Tasks was the strongest predictor of job performance. It is proposed that if measures were consistent there may have been a stronger effect of Multi-Tasks shown between the studies.

Study 2 suffered most in this regard because the way job performance ratings were gathered was not consistent between different business units, as various divisions had slightly different frameworks.

In addition to the limitations of the job performance measures, there are a number of other challenges that were faced, which were associated with performing studies with large groups of individuals across a number of organisations, both in Australia and cross-culturally. These challenges need to be considered when interpreting the results.

5.8 Idiosyncrasies of datasets and other challenges

It is also possible that the quality of the datasets may have affected study results. In the pilot study, study1 (Part B) and Study 3 invigilation of tests was very rigorous and it is believed that the tests were completed as per instructions. In Study 2, data was collected on a national basis by Human Resource Officers at the workplaces involved in the study and the author could not be present to monitor the quality of invigilation at all times. It is possible, however unverified, that invigilation may have been lenient at times and respondents may have had the opportunity to record notes using paper and pencil or other aids in order to improve their performance on the tests. This may explain the skewed nature on the Multi-Tasks and Digit Span tests. It was interesting to note that the other cognitive ability tests did not demonstrate the same pronounced effects. Using aids such as paper and pencil would benefit performance more so on counting-centric tasks.

Discrepancies in the research findings can also be attributed to a number of differences between the samples and other features of the studies. Firstly, the sample size in Study 3 was approximately one fifth the size of the sample in Study 1 and one tenth that of Study 2. Further, the Study 3 sample was less highly educated than in the previous studies. In Study 1 82% of the sample had a TAFE qualification or higher and almost half of those in Study 2 were educated to this level or above, whereas only 34% of Study 3 sample had greater than a high school education. The field in which the respondents worked in Study 3 (Manufacturing) was also very different to the white collar oriented roles that participants were applying for in Study 1 and 2. These differences may also explain the lower scores obtained on the intelligence measures in Study 3. Additionally, the job performance measures differed between the studies and job performance measures were not able to be collected for all respondents.

Lastly, the cross-cultural validity of the Multi-Tasks test was assessed by conducting research on this task and comparing performance with job performance indicators. This is a pioneering area of research and initial findings are discussed below.

5.9 Cross Cultural Considerations

Study 3 was the first of its kind to investigate the cross-cultural validity of the Multi-Tasks test. The results showed the test to be a valid and reliable measure within the Chinese culture, with outcomes that were comparable to those of the Australian participants in the previous studies. The factor structure of Multi-Tasks replicated that of Study 2 and differed from Study 1 only because the speed component was removed. Speed was not found to add value to the analysis in Study 1 and was not included in the subsequent studies. One of the models of intelligence that was found to fit the data in the present study replicated the model that emerged in Study 1. There was a relationship between Multi-Tasks and some indicators of job performance, though the relationship was not as strong as in previous studies, this was thought to be due to sampling differences and differences in the job performance measures used, rather than cultural differences. The lower scores on the intellectual functioning measures can also be explained by lower education levels and the ‘blue-collar’ careers of these participants, compared to the more highly educated ‘white-collar’ participants in the previous studies.

The representativeness of the sample also needs to be considered when interpreting the results. Over 70% of the participants were males and 66% were aged less than 35 years. Including a more varied group of workers, particularly older workers, may have strengthened the relationship between Multi-Tasks and job performance indicators. It is recommended that this pioneering study be replicated with further research on a larger sample of Chinese workers that are more representative of the working population. Certainly, Study 1 which included older workers and similar number of males and females showed a strong relationship between Multi-Tasks and two out of three job performance measures.

5.10 Conclusion

In conclusion, four studies have found the Multi-Tasks test, which comprises Placement Keeping (keeping track of ball movements) and Word Recall (recalling the order of words) sub-tasks to be a highly reliable measure. Its reliability exceeds that of two measures of Gc, three measures of Gf, one Gsm measure and five personality scales in an overall sample size of more than 2,500 workers in various industries across Australia and in China. It has been shown to be strongly predictive of supervisor ratings of job performance in an Australian context and some objective indicators of job performance both in Australia and China, particularly Job Level. Performance on the task has also been shown to be related to age and level of education, with better performance for older workers (with maximal performance between the age of 40-44) and those at higher levels within their organisation. These findings support previous research showing competing tasks paradigms to be particularly useful in predicting performance on complex jobs and in managerial roles (Ben-Shakhar & Sheffer, 2001; North & Gopher, 1976; Sauer et al., 1999; Stankov et al., 1989). There was also an effect for gender, but this was not consistent between studies. In terms of its place within the model of intelligence, based on the CHC Theory of Cognitive Abilities (McGrew, 2004; 2009), whilst it varies based on the number of measures included in the analysis and the cultural setting, overall Multi-Tasks has been found to be most strongly correlated with measures of Gf and Gsm. These are generally non-verbal and less influenced by culture or schooling than Gc measures. This also supports previous research on competing tasks measures (e.g., König, et al., 2005; Stankov, 1988). Measures of Gc that have a strong reasoning component and draw on working memory are also correlated with Multi-Tasks. As a result, Multi-Tasks has been shown to be a valid measure cross-culturally, with its factor structure, place within the intelligence model and reliability coefficients in the Chinese study replicating those of the previous studies. This pioneering group of studies shows Multi-Tasks to be a unique measure which is likely to be useful in predicting job performance and managerial potential in a range of jobs both in Australia and cross-culturally.

References

Adcock, R. A., Constable, R. T., Gore, J. C., & Goldman-Rakic, P. S. (2000). Functional neuroanatomy of executive processes involved in dual-task performance. *Proceedings of the National Academy of Sciences, USA, 97,* 3567-3572.

Anderson, J.R., Taatgen, N.A., & Byrne, M.D. (2005). Learning to achieve perfect timesharing: Architectural implications of Hazeltine, Teague, and Ivry (2002). Journal of Experimental Psychology: Human Perception and Performance, 31, 749-761.

Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive functions in monkeys. *Archives of General Psychiatry, 55,* 362-368.

Baddeley, A.D. (1990). *Human memory: Theory and practice*. Boston: Allyn and Bacon.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and performance: A meta-analysis. *Personnel Psychology, 44(1),* 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9(1-2),* 9-30. doi:10.1111/1468-2389.00160

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998).Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83(3),* 377-391. doi:10.1037/0021/9010.83.3.377

Bartram, D. (2004). Assessment in organisations. *Applied Psychology, 53(2),* 237-259.

Belsky, J. K. (1990). The psychology of aging theory, research, and interventions. Pacific Grove, CA: Brooks/Cole.

Ben-Shakhar, G., & Sheffer, L. (2001). The relationship between the ability to divide attention and standard measures of general cognitive abilities. *Intelligence, 29*, 293-306.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*, 410-424. doi:10.1037/0021-9010.92.2.410

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78*, 387-409.

Birkeland, S.A., Manson, T.M., Kisamore, J.L., Brannick, M.T., & Smith, M.A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317-335

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55*, 803-832.

Brown, S. W. (1998). Automaticity versus timesharing in timing and tracking dual-task performance. *Psychological Research, 61*, 71-81.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Beverly Hills, CA: Sage.

Brünken, R., Steinbacher, S., Plass, J., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual -task methodology. *Experimental Psychology, 42*(2), 109-119.

Buchweitz, A., Keller, T. A., Meyler, A., & Just, M. A. (in press). Brain activation for language dual-tasking: Listening to two people speak at the same time and a change in network timing. *Human Brain Mapping*. doi: 10.1002/hbm.21327

Bugg, J. M., Zook, N. A., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2006). Age differences in fluid intelligence: Contributions of general slowing and frontal decline. *Brain and Cognition, 62*(1), 9-16.

Bunge, S. A., Klingberg, T., Jacobsen, R. B., & Gabrieli, J. D. E. (2000). A resource model of the neural basis of executive working memory. *Proceedings of the National Academy of Sciences, USA, 97*, 3573-3578.

Buss, D. M. (1992). Manipulation in close relationships: Five personality factors in interactional context. *Journal of Personality, 60*, 477-499. doi:10.1111/j.1467-6494.1992.tb00981.x

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* New Jersey: Lawrence Erlbaum Associates.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*(2), 299-315. doi:10.1037/0096-3445.130.2.299

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

Carretta, T. R. (1987). *Time-sharing ability as a predictor of flight training performance.* Interim Technical Paper. Brooks Air Force Base, Texas.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122-130). New York: Guilford.

Cattell, R. B. (1971). *Abilities – their structure, growth, and action*. Boston: Houghton Mifflin.

Chan, A. S., Shum, D., & Cheung, R. W. Y. (2003). Recent development of cognitive and neuropsychological assessment in Asian countries. *Psychological Assessment, 15*(3), 257-267. doi:10.1037/1040-3590.15.3.257

Chou, K. L., & Chi, I. (2000). Assessing Chinese adults' intellectual abilities. *Clinical Gerontologist, 21*(3), 65-72.

Chisholm, C. D., Dornfeld, A. M., Nelson, D. R., & Cordell, W. H. (2001). Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine*, 38(2), 146-151.

Collette, F., Olivier, L., Van der Linden, M., Laureys, S., Delfiore, G., Luxen, A., et al. (2005). Involvement of both prefrontal and inferior parietal cortex in dual-task performance. *Cognitive Brain Research, 24,* 237-251.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P.C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32,* 277-296.

Colonia-Willner, R. (1998). Practical intelligence at work: Relationship between aging and cognitive efficiency among managers in a bank environment. *Psychology and Aging, 13*(1), 45-57. doi: http://dx.doi.org/10.1037/0882-7974.13.1.45

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163-183.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness. *Personality & Individual Differences 12*, 887-898.

Costa, P.T., Jr., & McCrae, R.R. (1992). *Revised NEO personality inventory and NEO five factor inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378,* 279-281.

Damos, D., & Smist, T. (1982). Individual differences in multi-task response strategies. *Aviation, Space and Environmental Medicine, 53*(12), 1177-1181.

Dattel, A. R., & King, R. E. (2010). The impact of reweighting the AT-SAT subtests on group score differences. *The International Journal of Aviation Psychology, 20*(2), 111-123.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica, 101*, 395-416.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*(1), 417-440.

Digman, J. M. (1997). Higher order factors of the Big Five. *Journal of Personality & Social Psychology, 73*, 1246-1256.

Ekstrom, R. M., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs,79*(2), 3-84.

Ellenbogen, R., & Meiran, N. (2011). Objects and events as determinants of parallel processing in dual tasks: Evidence from the backward compatibility effect. Journal of Experimental Psychology: *Human Perception and Performance, 37*(1), 152-167.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Eysenck, H. J., & Eysenck, M.W. (1985). *Personality and individual differences*. New York: Plenum.

Fay, D., & Frese, M. (2001). The concept of Personal Initiative: An overview of validity studies. *Human Performance, 14*(1), 97-124.

Fan, W., Zhang, L-F., & Watkins, D. (2010). Incremental validity of thinking styles in predicting academic achievements. *Educational Psychology, 30*, 605-623. doi: 10.1080/01443410.2010.496899

Fischer, H., Wik, G., & Fredrikson, M. (1997). Extraversion, neuroticism and brain function: A PET study of personality. Personality and Individual Differences, 23, 345-352.

Fogarty, G. (1987). Timesharing in relation to broad ability domains. *Intelligence*, *3,* 207-231.

Fogarty, G., & Stankov, L. (1982). Competing tasks as an index of intelligence. *Personality and Individual Differences, 3,* 407-422.

Fogarty, G., & Stankov, L. (1995). Challenging the "Law of diminishing returns". *Intelligence*, *21*, 157-174.

Ghiselli, E. E. (1963). Intelligence and managerial success. *Psychological Reports*, *12,* 898-898.

Gill, R. (1982). A trainability concept for management potential and an empirical study of its relationship with intelligence for two managerial skills. *Journal of Occupational Psychology*, *55*, 139-147.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. doi:10.1037/0003-066X.48.1.26

Gonzalez, C., Thomas, R.P., & Vanyukov, P. (2005).The relationship between cognitive abilityand dynamic decision making. *Intelligence, 33*, 169-186.

Goodstein, L. D., & Lanyon, R. I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business and Psychology, 13*(3), 291-322. doi:10.1023/A:1022941331649

Gong, Y. X., Zhang, Z. X., Cheng, Z. H., Cai, T. S., Meng, X .Z., Zhou, S. J., et al. (1995). Norms of the nonverbal intelligence tests on the minority nationality. *Chinese Journal of Clinical Psychology, 3*, 129–136. (in Chinese)

Goonetilleke, R. S., & Luximon, Y. (2010). The relationship between monochronicity, polychronicity and individual characteristics. *Behaviour & Information Technology, 29*(2), 187-198.

Gottfredson, L. S. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior, 29*, 376-410.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*(2), 135-164.

Greenham, S. L., Stelmack, R. M., & Campbell, K. B. (2000). Effects of attention and semantic relation on event-related potentials in a picture-word naming task. *Biological Psychology, 50,* 79-104.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. L. (2010).*Multivariate data analysis, 7thed*. New Jersey: Prentice-Hall.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology, 24*(8), 1149-1167. doi: 10.1002/acp.1624.

Hambrick, D.Z., Rench, T.A., Poposki, E.M., Darowski, E.S., Roland, D., Bearden, R.M., Oswald, F.L., &Brou, R. (2011). The relationship between the ASVAB and multitasking in navy sailors: a process-specific approach. *Military Psychology, 23*(4), 365-380.

Higgins, L. T., & Xiang, G. (2009). The development and use of intelligence tests in China. *Psychology & Developing Societies, 21*, 257-275. doi: 10.1177/097133360902100205

Hofheinz, M., & Schusterschitz, C. (2010). Dual task interference in estimating the risk of falls and measuring change: A comparative, psychometric study of four measurements. *Clinical Rehabilitation, 24*, 831-842.

Hogan, R. (2005). In defence of personality measurement. *Human Performance, 18*(4), 331-341.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-577. doi:10.1037/0003-066X.51.5.469

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job performance relations. *Journal of Applied Psychology, 88*, 100-112. doi:10.1037/0021-9010.88.1.100

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive energetical framework. *Biological Psychology, 45*, 73-93.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories*, *tests and issues* (pp. 53-91). New York: Guilford.

Howell, D. C. (1992). *Statistical methods for psychology*. Boston: PWS-Kent.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal, 6*, 1-55.

Hunt, E. (1980). Intelligence as an information-processing concept. *British Journal of Psychology, 71*, 449-474.

Hunter, J. E., & Hunter, R. F. (1984).Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hunter, J., Schmidt, F., & Judiesch, M. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*(1), 28-42.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85,* 869-879. doi:10.1037/0021-9010.85.6.869

Irwin-Chase, H., & Burns, B. (2000). Developmental changes in children's abilities to share and allocate attention in a dual task. *Journal of Experimental Child Psychology, 77*, 61-85.

Jang, K. L., Livesley, W. J., & Vemon, P. A. (1996). Heritability of the Big Five personality dimensions and their facets: A twin study. *Journal of Personality, 64,* 577-592. doi:10.1111/j.1467-6494.1996.tb00522.x

Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., et al. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in singe and dual tasks: Functional magnetic resonance imaging study. *Neuroimage, 19,* 210-225.

Janssen, C. P., & Brumby, D. P. (2010).Strategic adaptation to performance objectives in a dual-task setting. *Cognitive Science, 34*, 1548-1560.

Jensen, A. R. (1998). The g factor: The science of mental abilities. Westport, CT: Praeger.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control and perception of relationships in the environment. *Journal of Personality and Social Psychology, 78*, 943-955.

Jing, Q. (1994). Development of psychology in China. *International Journal of Psychology*, *29*, 667-675.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability and career success across the life span. *Personnel Psychology*, *52*, 621-652.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*, 797-807. doi:10.1037/0021-9010.87.4.797

Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L., Zajac, H., & Thulborn, K. R. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *NeuroImage*, *14*, 417-426.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J., Prentice Hall.

Kail, R. (1991). Developmental changes in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490-501.

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

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47-70.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66-71.

Kaufman, A. S., Johnson, C. K., & Liu, X. (2008). A CHC Theory-Based analysis of age differences on cognitive abilities and academic skills at ages 22 to 90 years. *Journal of Psychoeducational Assessment, 26*, 350-381.

Ketelsen, K., & Welsh, M. (2010).Working memory and mental arithmetic: a case for dual central executive resources. *Brain and Cognition*, *74*(3), 203-209.

King, R. E., Manning, C. A., & Drechsler, G. K. (2006). Operational use of the Air Traffic Selection and Training Battery. *International Journal of Applied Aviation Studies, 6*(2), 207-217.

Klingberg, T. (1998). Concurrent performance of two working memory task: Potential mechanisms of interference. *Cerebral Cortex, 8*, 593-601.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). NewYork: Guilford.

Kolb, D. (1985). *Learning Styles Inventory*. Boston: McBer.

Kondo, H., Osaka, N., & Osaka, M. (2004). Cooperation of the anterior cingulate cortex and dorsolateral prefrontal cortex for attention shifting. *Neuroimage*, *23*, 670-679.

König, C. J., Bühner, M., & Mürling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multi-tasking performance, but polychronicity and extraversion are not. *Human Performance, 18*(3), 243-266.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 35-60). New York: Academic Press.

Kramer, A. F., Wickens, C., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 393-408.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: a test of the four sources model. *The Journal of General Psychology, 120*, 375-405.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389-433.

Lam, S. S. K. (1998). Organizational performance and learning styles in Hong Kong. *The Journal of Social Psychology, 138*(3), 401-402.

Le, H., Oh, I., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology, 96*(2), 113-133. doi:10.1037/a0021016

Le Moal, M., & Simon, H. (1991). Mesocorticolimbic dopaminergic network: Functional and regulatory roles. *Physiological Reviews, 71*,155-234.

Lieberman, M. D. (2000). Introversion and working memory: central executive differences. *Personality and Individual Differences, 28*, 479-486.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. S*tructural Equation Modeling, 9*(2), 151-173.

Liu, X. F., Miao, D. M., Wang, G. X., An, C., Pan, C. B., Wang, W., et al. (2001). Reliability and validity analysis of group intelligence test of army academy students. *Journal of Fourth Military Medicine University, 19*, 1821–1824. (in Chinese)

Little, T. D., Cunningham, W.A., Shahar, G., & Widaman, K.F. (2002). To parcel or not to parcel: exploring the question, weighing the merits. *Structural Equation Modeling, 9*(2), 151-173.

Lounsbury, J. W., Loveland, J. M., Sundstrom, E. D., Gibson, L. W., Drost, A. W., & Hamrick, F. L. (2003). An investigation of personality traits in relation to career satisfaction. *Journal of Career Assessment, 11*(3), 287-307. doi:10.1177/1069072703254501

Lorsbach, T., & Simpson, G. (1988). Dual -task performance as a function of adult age and task complexity. *Psychology and Aging, 3*(2), 210-212.

Luer, G., Becker, D., Lass, U., Yunqiu, C., & Zhongming, W. (1998). Memory span in German and Chinese: Evidence for the phonological loop. *European Psychologist, 3*(2), 102-112.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32,* 481-498.

Marr, D. B., & Sternberg, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and nongifted children to relevant and irrelevant novel stimuli. *Cognitive Development, 1*, 53-72.

Martin, J. H. (1996). *Neuroanatomy: Text and atlas* (2nd ed.). Stamford, CT: Appleton & Lange.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*(4), 493-505. doi: 10.1037/0021-9010.79.4.493

McCrae, R. R., & Costa, P. T. (1991). The NEO Personality Inventory: Using the Five-Factor Model in counseling. *Journal of Counseling & Development, 69*(4), 367-372.

McCrae, R. R., & Costa, P. T. (2008). Empirical and theoretical status of the five-factor model of personality traits. In J. Boyle. G. Matthew, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (pp. 273- 294). Los Angeles: Sage.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel *Psychology*, *43*, 335-354.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can we trust the data? *Psychological Science*, *11*, 502-506.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). New York: Guilford.

McGrew, K. S. (2004). Cattell-Horn-Carroll theory of cognitive abilities CHC (Gf-Gc): Past, Present & Future Retrieved 18 April, 2007, from http://www.iapsych.com/CHCPP/CHCPP.html

McGrew, K. S. (2005). *The Cattell-Horn-Carroll theory of cognitive abilities*. New York: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1-10.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual Woodcock- Johnson III*. Itasca, IL: Riverside Publishing.

McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology, 79*, 836-844. doi: 10.1037/0021-9010.79.6.836

Mok, K-H. (2005). Globalization and educational restructuring: University merging and changing governance in China. *Higher Education, 50*, 57-88. doi: 10.1007/s10734-004-6347-z

Mount, M. K., & Barrick, M. R. (1998). Five reasons why the big five article has been frequently cited. *Personnel Psychology*, *51*(4), 849-857. doi: 10.1111/j.1744-6570.1998.tb00743.x

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11(2/3)*, 145-165. doi: 10.1207/s15327043hup1102 & 3_3

Mount, M. K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, *61*(1), 113-139. doi:10.1111/j.1744-6570.2008.00107.x

Miller, J. O., Ulrich, R., & Rolke, B. (2009). On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies. *Cognitive Psychology, 58*(3), 273-310.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Mousavi, S. M., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

Murray, C. (1998). *Income and inequality*.Washington, DC: AEI Press.

Nelton, S. (1994). Put your purpose in writing. *Nation's Business, 82*, 61-64.

Nettelbeck, T. (1994). Speediness. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 1014-1019). New York: Macmillan.

Nettelbeck, T., & Rabbitt, P.M. (1992). Aging, cognitive performance, and mental speed. *Intelligence, 16*(2), 189-205.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist, 61*(6), 622-631. doi:10.1037/0003-066X.61.6.622

Newman, S. D., Keller, T. A., & Just, M. A. (2007). Volitional control of attention and brain activation in dual task performance. *Human Brain Mapping, 28*, 109-117.

North, R., & Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors, 18*(1), 1-14.

Northcraft, G. B., Schmidt, A. M., & Ashford, S. (2011). Feedback and the Rationing of Time and Effort Among Competing Tasks. *Journal of Applied Psychology, Advance online publication* doi: 10.1037/a0023221

Nyborg, H. (2003). *The scientific study of general intelligence: Tribute to Arthur R. Jensen.* San Diego: Pergamon.

Oberauer, K., & Göthe, K. (2006). Dual-task effects in working memory: Interference between two processing tasks, between two memory demands, and between storage and processing. *European Journal of Cognitive Psychology, 18*(4), 493-519.

Oh, I-S., Wang, G., & Mount, M. K. (2010). Validity of observer ratings of the Five-Factor Model of Personality Traits: A meta-analysis. *Journal of Applied Psychology, Advance online publication*. doi:10.1037/a0021832

Omnibus Screening Protocol. (2001). Version 1.3 & 1.4. October & December, 2001.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*(4), 995-1027. doi:10.1111/j.1744-6570.2007.00099.x

O'Reilly, C. A., & Chatman, J. A. (1994). Working smarter and harder: A longitudinal study of managerial success. *Administrative Science Quarterly, 39*, 603-627.

Oberauer, K., & Göthe, K. (2006). Dual-task effects in working memory: Interference between two processing tasks, between two memory demands, and between storage and processing. *European Journal of Cognitive Psychology, 18*(4), 493-519.

Ozer, D. J., & Reise, S. P. (1994). Personality assessment. *Annual Review of Psychology, 45*(1), 357-388.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt, Rinehart and Winston.

Rammsayer, T. H. (1998). Extraversion and dopamine: Individual differences in response to changes in dopaminergic activity as a possible biological basis of extraversion. *European Psychologist, 3*(1), 37-50.

Rammsayer, X., Netter, P., & Vogel., W. H. (1993). A neurochemical model underlying differences in reaction times between introverts and extraverts. *Personality and Individual Differences, 14,* 701-712.

Raven, J., Raven, J.C., & Court, J.H. (1998). *Raven manual: Section 4, advanced progressive matrices, 1998 edition*. Oxford, UK: Oxford Psychologists Press Ltd.

Resnick, S. M., Goldszal, A. F., Davatzikos, C., et al. (2000). One-year age changes in MRI brain volumes in older adults. *Cerebral Cortex, 10,* 464-472.

Resnick, S. M., Pham, D. L., Kraut, M. A., Zonderman, A. B., & Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain. *Journal of Neuroscience, 23,* 3295–3301.

Roberts, R., Beh, H., & Stankov, L. (1988). Hick's law, competing-task performance, and intelligence. *Intelligence, 12,* 111-130.

Roberts, R., Beh, H., Spilsbury, G., & Stankov, L. (1991). Evidence for an attentional model of human intelligence using the competing task paradigm. *Personality and Individual Differences, 12*(5), 445-455.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644. doi: 10.1037/0021-9010.83.4.634

Rushton, J. P., Skuy, M., & Bon, T. A. (2004). Construct validity of Raven's Advanced Progressive Matrices for African and non-African engineering students in South Africa. *International Journal of Selection & Assessment, 12*, 220-229.

Ruthruff, E., Van Selst, M., Johnston, J. C., & Remington, R. (2006). How does practice reduce dual-task interference: Integration, automatization or just stage-shortening? *Psychological Research, 70*, 125-142.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88*, 1068-1081.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition, 30*, 572-582.

Sauer, J., Wastell, D. G., & Hockey, G. R. J. (1999).Multiple-task performance on a computer-simulated life support system during a space mission simulation. *Acta Astronautica, 44*(1), 43-52.

Schmidt, A. M., & DeShon, R. P. (2007). What to do? The effects of discrepancies, incentives, and time on dynamic goal prioritization. *Journal of Applied Psychology, 92*, 928 -941.

Schmidt, F. L., & Hunter, J. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science, 1*(3), 89-92.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274. doi: 10.1037/0033-2909.124.2.262

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162-173. doi:10.1037/0022-3514.86.1.162

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*(3), 407-422.

Schmitt, N., & Mills, A. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology, 86*(6), 451-458.

Schulze, R., & Roberts, R.D. (2006). Assessing the Big Five: Development and validation of the Openness Conscientiousness Extraversion Agreeablness Neuroticism Index Condensed. *Zeitschrift fur Psychologie, 214*(3), 133-149.

Schumacher, E. H., & Schwarb, H. (2009). Parallel response selection disrupts sequence learning under dual-task conditions. *Journal of Experimental Psychology: General, 138*(2), 270-290.

Seibert, S. E., & Kraimer, M. L. (2001).The five-factor model of personality and career success. *Journal of Vocational Behavior*, *58*(1), 1–21. doi:10.1006/jvbe.2000.1757

Shaver, P. R., & Brennan, K. A. (1992). Attachment styles and the “big five” personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin, 18*, 536-545. doi: 10.1177/0146167292185003

Shirey, L. L., & Reynolds, R. E. (1988). Effects of interest on attention and learning. *Journal of Educational Psychology, 80*(2), 159-166.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*, 215-230.

Spilsbury, G., Stankov, L., & Roberts, R. (1990). The effect of a test's difficulty on its correlation with intelligence. *Personality and Individual Differences, 11*(10), 1069-1077.

Stankov, L. (1983). Attention and intelligence. *Journal of Educational Psychology, 75*(4), 471-490.

Stankov, L. (1983b). The role of competition in human abilities revealed through auditory tests. *Multivariate Behavioral Research Monographs, 83*(1), 63-VII.

Stankov, L. (1986). Age-related changes in auditory abilities and in a competing task. *Multivariate Behavioral Research, 21*, 65-76.

Stankov, L. (1988). Single tests, competing tasks and their relationship to the broad factors of intelligence. *Personality and Individual Differences*, *9*(1), 25-33.

Stankov, L. (1989). Attentional resources and intelligence: A disappearing link. *Personality and Individual Differences, 10,* 957-968.

Stankov, L. (1999). Capacity versus speed in intelligence and ageing. *Australian Psychologist*, *34*(2), 138-143.

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*(2), 121-143.

Stankov, L. (2003). Complexity in human intelligence. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of intelligence* (pp. 27-42). Washington, DC.: American Psychological Association.

Stankov, L., Fogarty, G., & Watt, C. (1989). Competing tasks: Predictors of managerial potential. *Personality and Individual Differences, 10*(3), 295-302.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology*, *72*(1), 21-44.

Stankov, L., & Myors, B. (1990). The relationship between working memory and intelligence: Regression and COSAN analyses. *Personality and Individual Differences*, *11*(10), 1059-1068.

Stankov, L., & Roberts, R. (1997). Mental speed is not the 'basic' process of intelligence. *Personality and Individual Differences, 22*(1), 69-84.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-component tests are not much more than g: An extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*(3), 193-205.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests* (3rd edition). Oxford: Oxford University Press

Sternberg, R. J. (1982). Natural, unnatural, and supernatural concepts. *Cognitive Psychology*, *14*, 451-488.

Sternberg, R. J. (1997). Managerial intelligence: Why IQ isn't enough. *Journal of Management, 23*, 475-493.

Szameitat, A. J., Schubert, T., Müller, K., & von Cramen, D. Y. (2002). Localization of executive function in dual-task performance with fMRI. *Journal of Cognitive Neuroscience 14*, 1184-1199.

U.S. Department of Education. (1997). *The Condition of Education*. Retrieved 7 April, 2011, from http://www.ed.gov/NCES

Varela, O.E., Salgado, E.I., & Lasio, M.V. (2010). The meaning of job performance in collectivistic and high power distance cultures. *Cross Cultural Management: An International Journal, 17*(4), 407-426.

Viswesvaran, C., & Ones, D. S. (1999) Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210.

Walsh, W. B., & Eggerth, D. E. (2005). Vocational psychology and personality: The relationship of the Five-Factor model to job performance and job satisfaction. In W. B. Walsh & M. L. Savickas (Eds.), *Handbook of vocational psychology: Theory, research, and practice* (3rd ed., pp. 267-295). New Jersey: Lawrence Erlbaum Associates.

Wang, D-Y. D., Proctor, R. W., & Pick, D. F. (2007). Acquistion and transfer of attention allocation strategies in a multiple-task work environment. *Human Factors, 49*(6), 995-1004.

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review, 17*, 479-485.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Wickens, C., & Seidler, K. (1997). Information access in a dual-task context: Testing a model of optimal strategy selection. *Journal of Experimental Psychology: Applied, 3*(3), 196-215.

Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence, 30*, 537-554.

Wilk, S. L., Desmarais, L. B., & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology, 80*, 79-85.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & Stress, 14*(1), 74-99.

Zhang, L-F. (1999). A comparison of U.S. and Chinese university student's cognitive development: the cross-cultural applicability of Perry's theory. *The Journal of Psychology, 133*, 425-439.

Zhou, S. S., Fan, J., Le, T. M. C., Wang, C., & Wang, K. (2011). Age-related differences in attentional networks of alerting and executive control in young, middle-aged, and older Chinese adults. *Brain and Cognition, 75*, 205-210.

Zimmerman, R. D., Triana, M. D. C., & Barrick, M. R. (2010). Predictive criterion-related validity of observer ratings of personality and job-related competencies using multiple raters and multiple performance criteria. *Human Performance, 23*(4), 361-378. doi: 10.1080/08959285.2010.501049

Appendix 1A

Histogram of Variables for Study 1 (Part B)

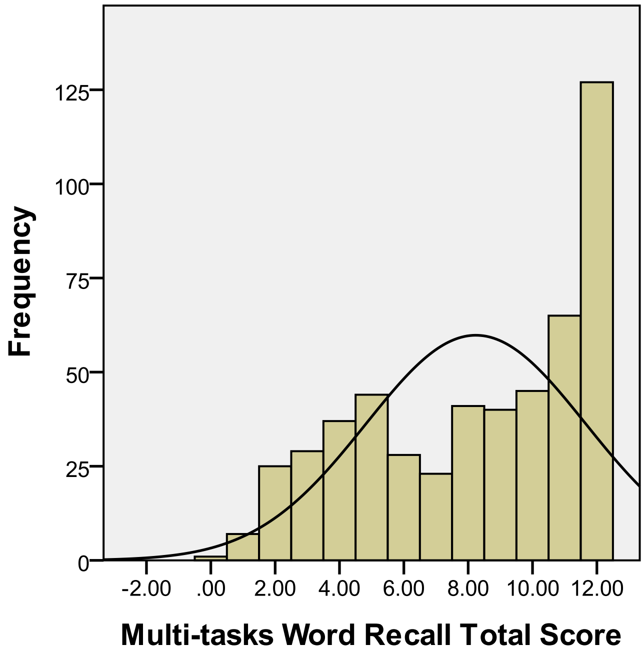


Figure 1.A1. Histogram for Multi-Tasks Word Recall total score.

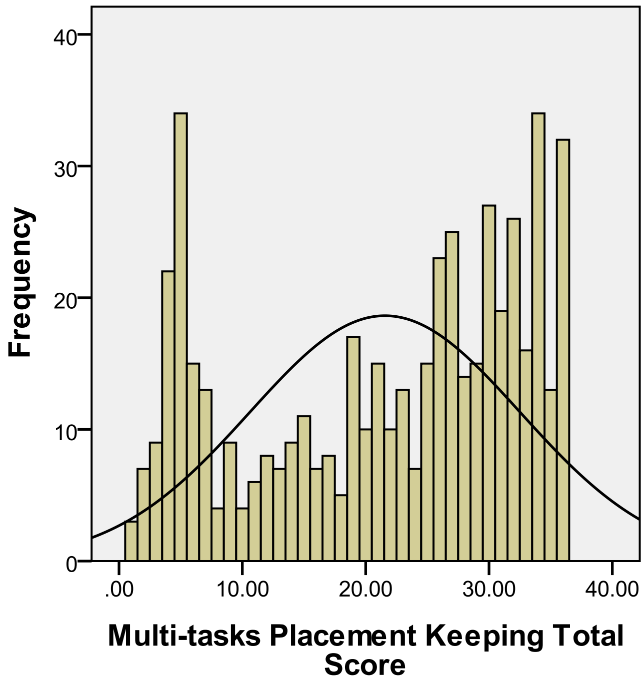


Figure 1.A2. Histogram for Multi-Tasks Placement Keeping total score.

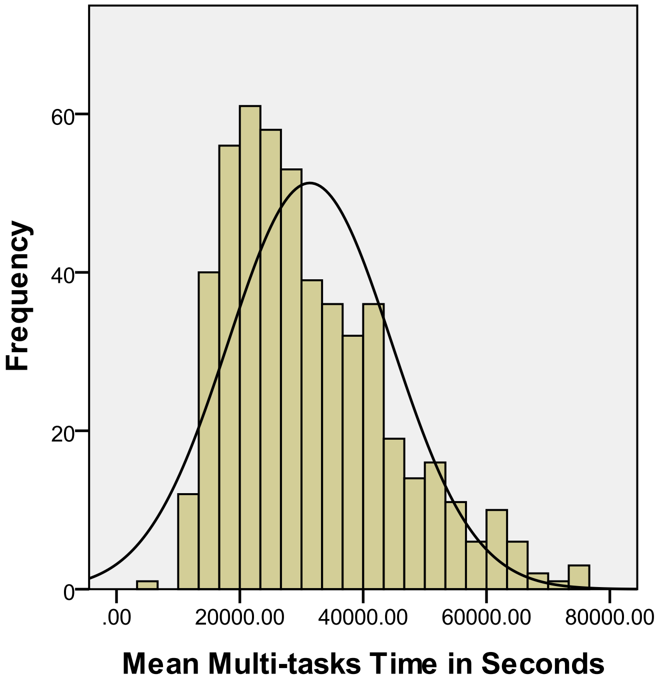


Figure 1.A3. Histogram for Multi-Tasks time in milliseconds.

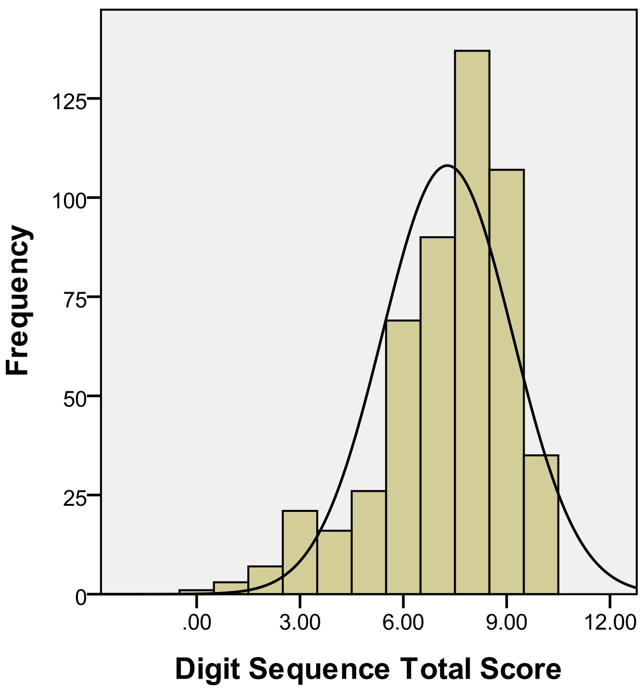


Figure 1.A4. Histogram for Digit Sequence total score.

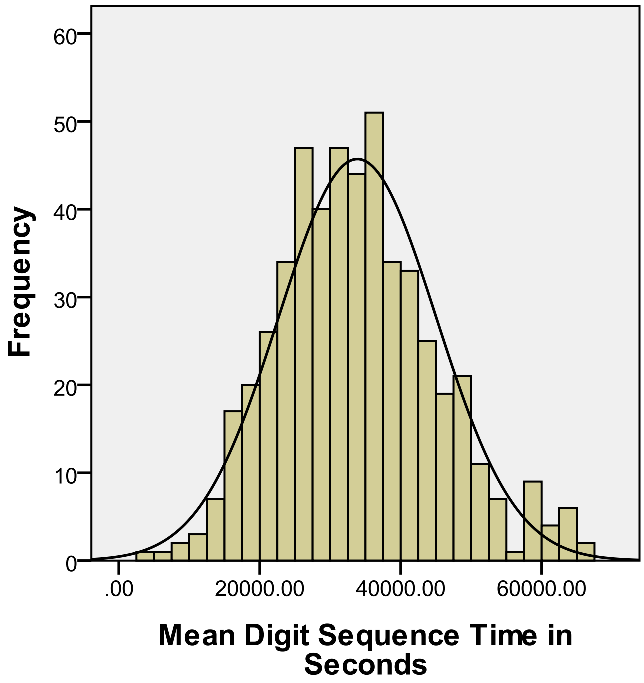


Figure 1.A5. Histogram for Digit Sequence time in milliseconds.

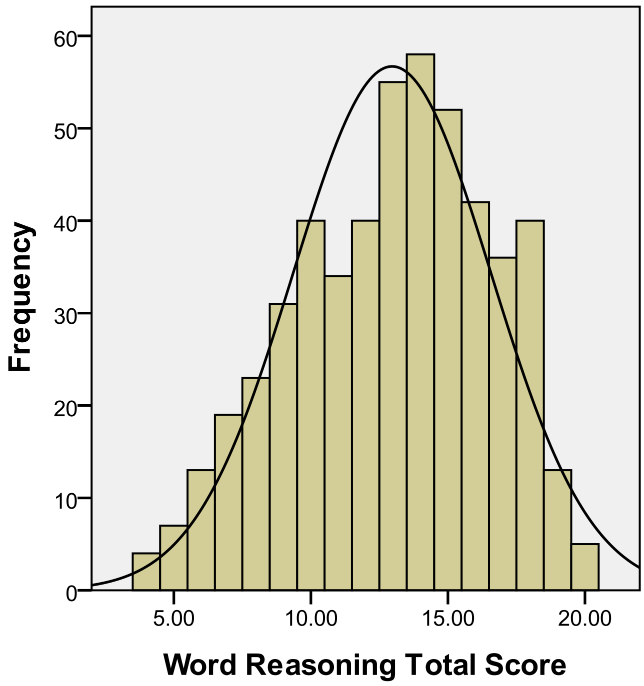


Figure 1.A6. Histogram for Word Reasoning total score.

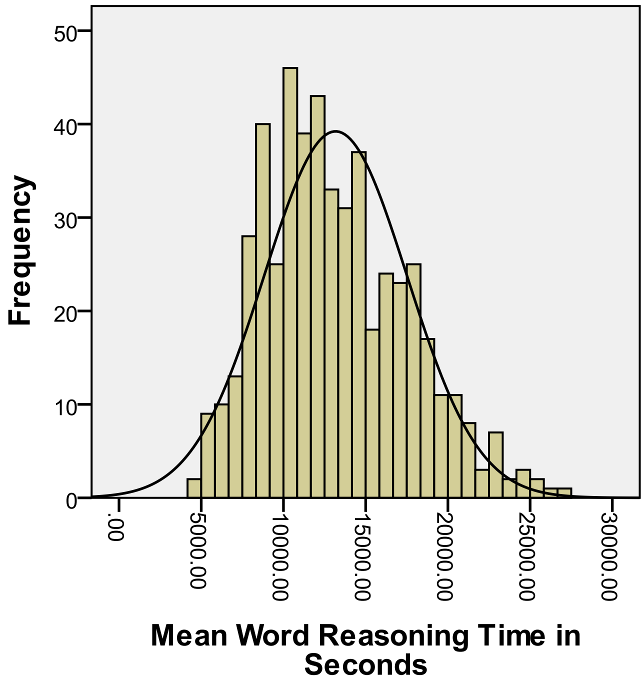


Figure 1.A7. Histogram for Word Reasoning time in milliseconds.

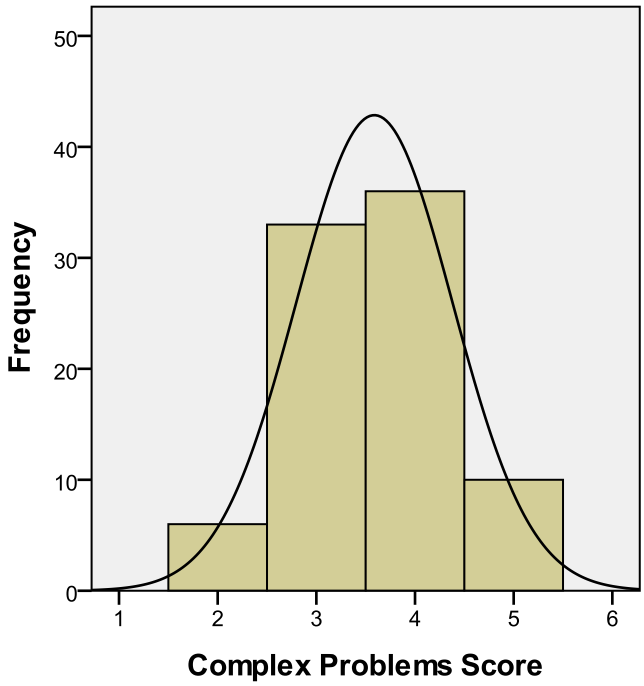


Figure 1.A8. Histogram for Complex Problems score.

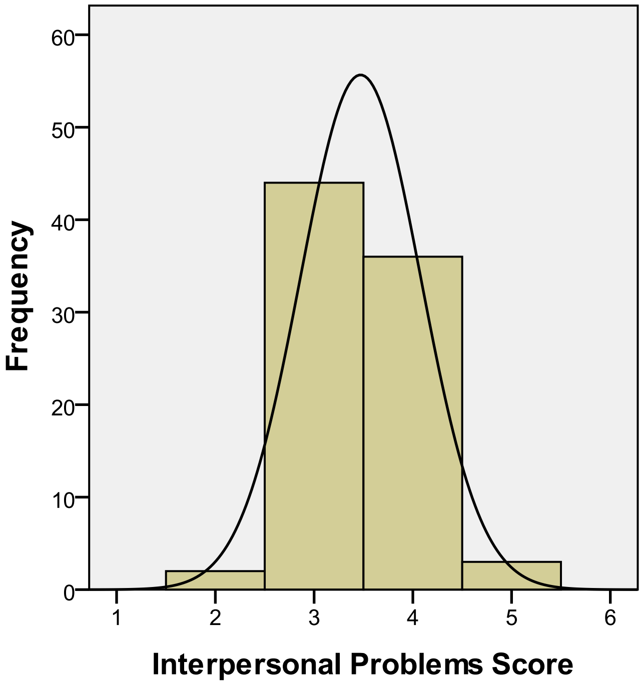


Figure 1.A9. Histogram for Interpersonal Problems score.

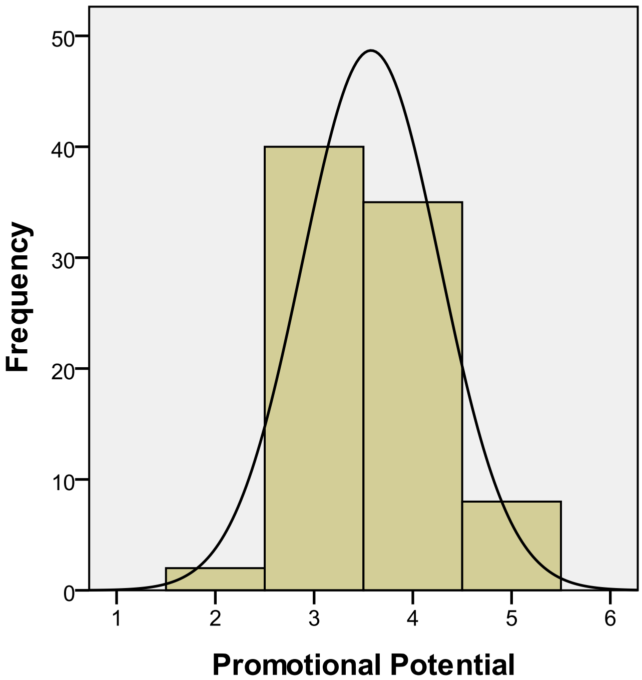


Figure 1.A10. Histogram for Promotional Potential score.

Appendix 1B

Composition of Parcels for Study 1 (Part B)

Parcels for Digit Sequence

1. Parcel 1 – 152, 167, and 182
2. Parcel 2 – 147 and 197
3. Parcel 3 – 157 and 192
4. Parcel 4 – 162, 177, and 202

Parcels for Word Reasoning

1. Parcel 1 – 403, 409, 410, 412, and 413
2. Parcel 2 – 394, 398, 399, and 404
3. Parcel 3 – 392, 401, and 414
4. Parcel 4 – 400, 405, 406, and 407
5. Parcel 5 – 390, 396, 408, and 411

Parcels for Multi-Tasks Tests

1. Parcel 1 – 2010, 2011, and 2012
2. Parcel 2 – 2013, 2014, and 2015
3. Parcel 3 – 2017, 2018, and 2020
4. Parcel 4 – 2021, 2022, and 2024

Appendix 1C

Confirmatory Factor Analysis Results for Study 1 (Part B)

Table 1.C1

Path coefficients for the Multi-Tasks measurement model

Path	B	SE	β	CR	Sig.
Multi-Tasks word recall to:					
Parcel 4	1.00		.80		
Parcel 3	3.21	.16	.84	19.27	.000
Parcel 2	3.03	.16	.79	18.35	.000
Parcel 1	1.95	.15	.58	12.80	.000
Multi-Tasks placement keeping to:					
Parcel 4	1.00		.83		
Parcel 3	10.31	.45	.84	22.64	.000
Parcel 2	1.06	.04	.91	25.19	.000
Parcel 1	.94	.04	.81	21.40	.000
Multi-Tasks broad speediness to:					
Parcel 4	1.00		.86		
Parcel 3	.95	.03	.90	26.31	.000
Parcel 2	.78	.03	.85	24.54	.000
Parcel 1	.61	.03	.71	18.53	.000

Table 1.C2

Co-variances between constructs for the Multi-Tasks measurement model

Relationship	B	SE	CR	Sig.
Multi-Tasks word recall and:				
Placement keeping	1.08	.14	7.76	.000
Broad speediness	14.60	2.14	6.80	.000
Multi-Tasks placement keeping and:				
Broad speediness	39.27	5.55	7.06	.000

Appendix 1D

Path Coefficients for the Intelligence Model for Study 1

Table 1.D1

Path coefficients for the intelligence model

Relationship	B	SE	β	CR	Sig.
Fluid intelligence to:					
Multi-tasks placement keeping	1.68	.24	.87	7.07	.000
Multi-tasks word recall	.30	.05	.42	6.05	.000
Digit sequence	1.00		.75		
Word reasoning (Gc) to:					
Word 1	7.23	.70	.61	10.33	.000
Word 2	7.68	.71	.66	10.88	.000
Word 3	.82	.08	.58	9.96	.000
Word 4	1.00		.63		
Word 5	4.75	.48	.58	10.00	.000

Appendix 2A

Histograms of Variables for Study 2

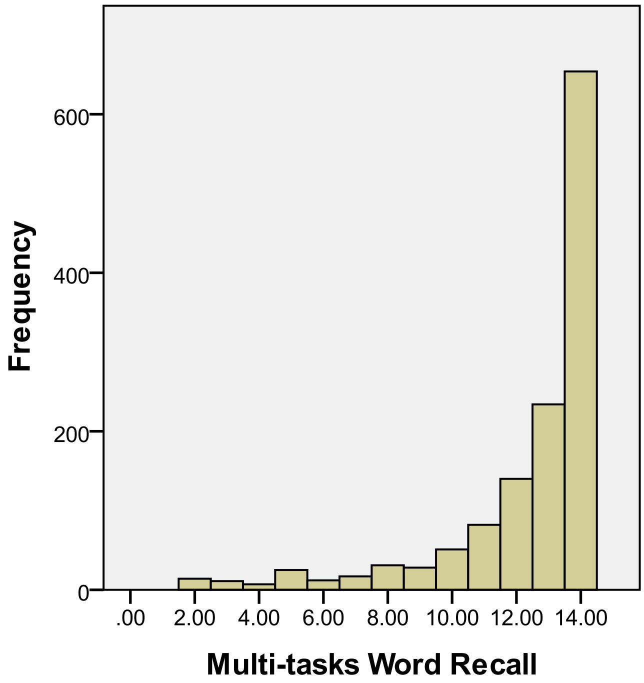


Figure 2.A1. Histogram for Multi-Tasks Word Recall total score.

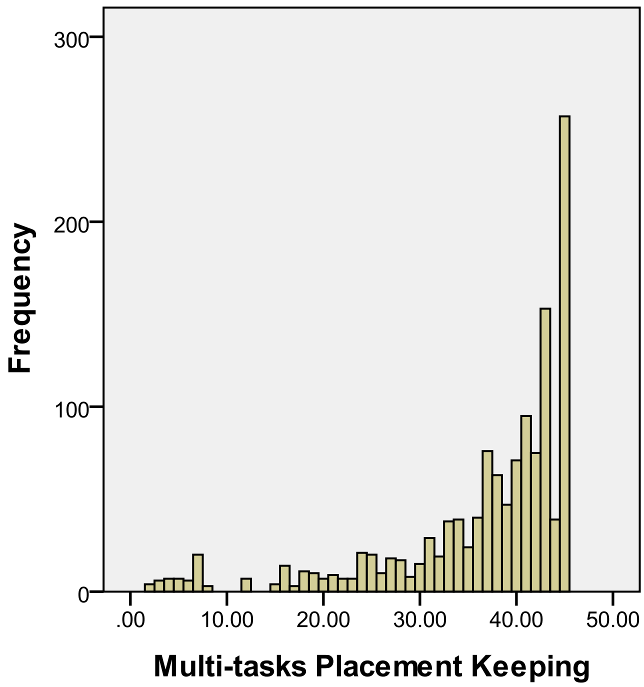


Figure 2.A2. Histogram for Multi-Tasks Placement Keeping total score.

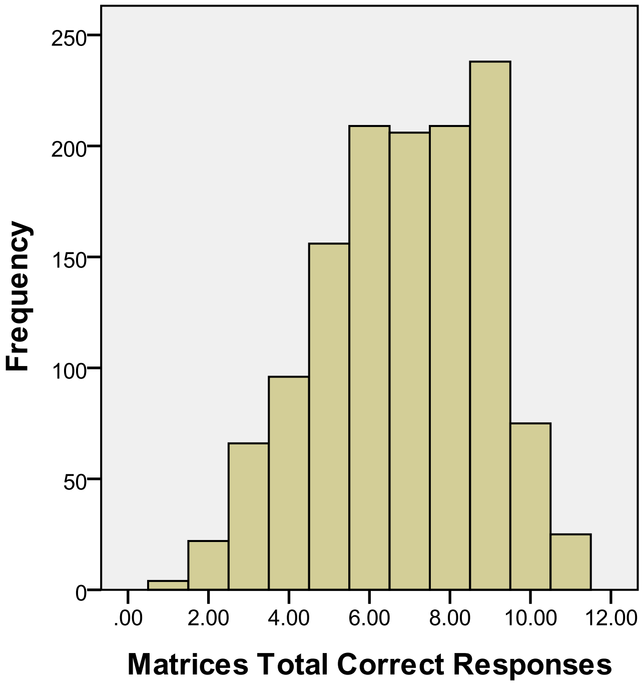


Figure 2.A3. Histogram for Matrices total score.

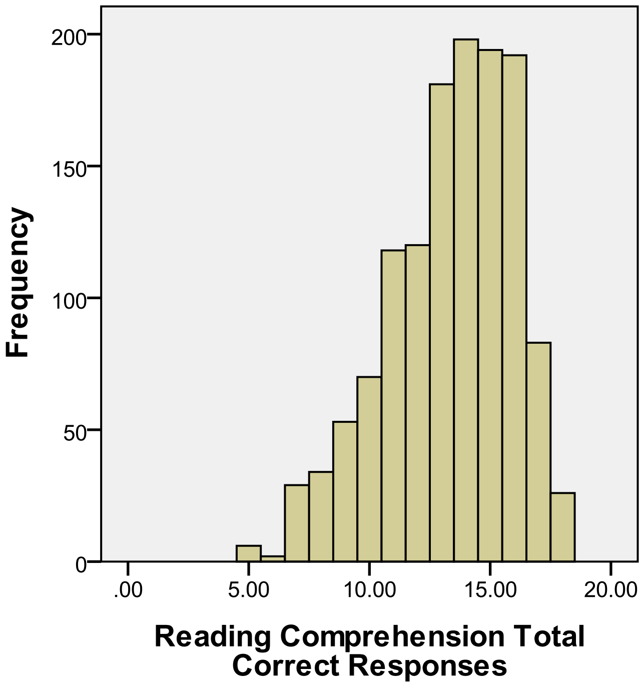


Figure 2.A4. Histogram for Reading Comprehension total score.

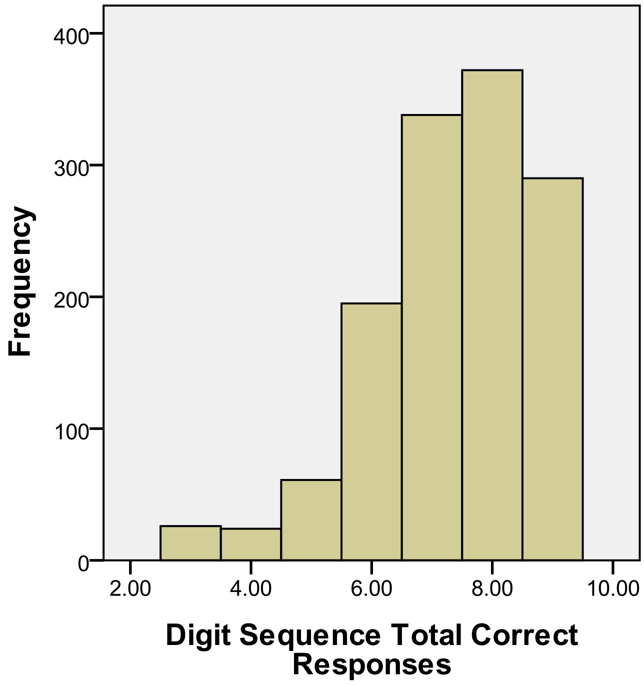


Figure 2.A5. Histogram for Digit Sequence total score.

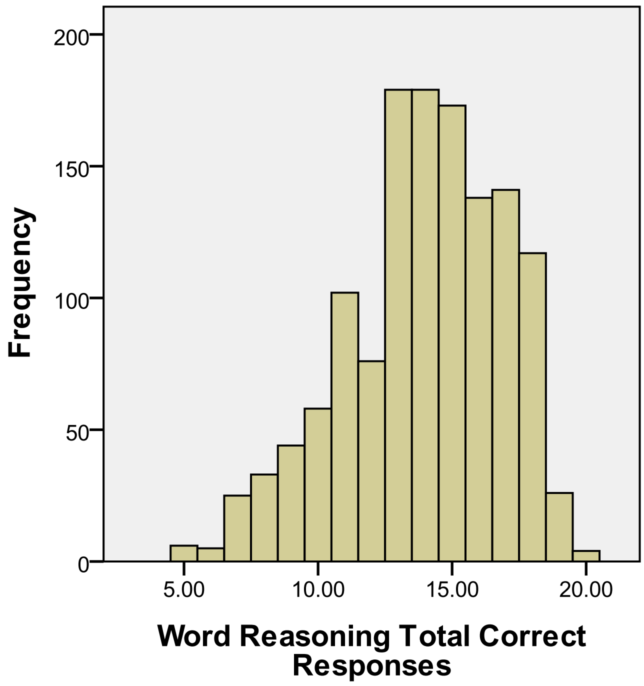


Figure 2.A6. Histogram for Word Reasoning total score.

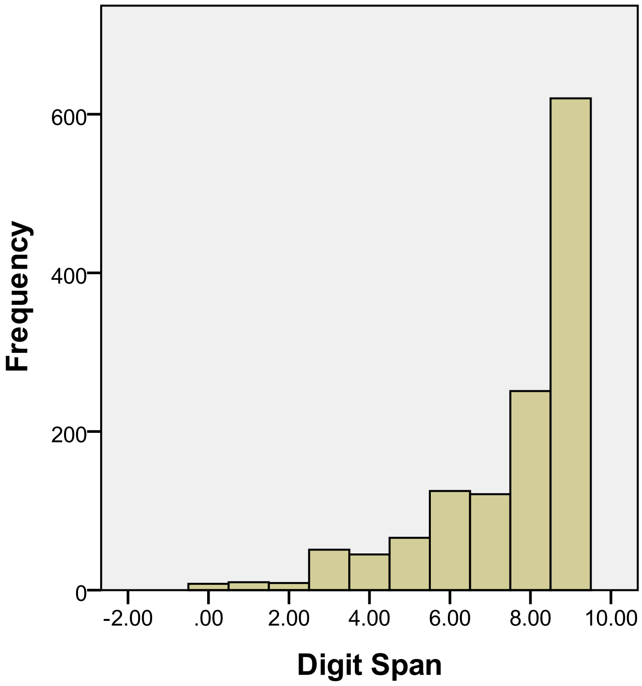


Figure 2.A7. Histogram for Digit Span total score.

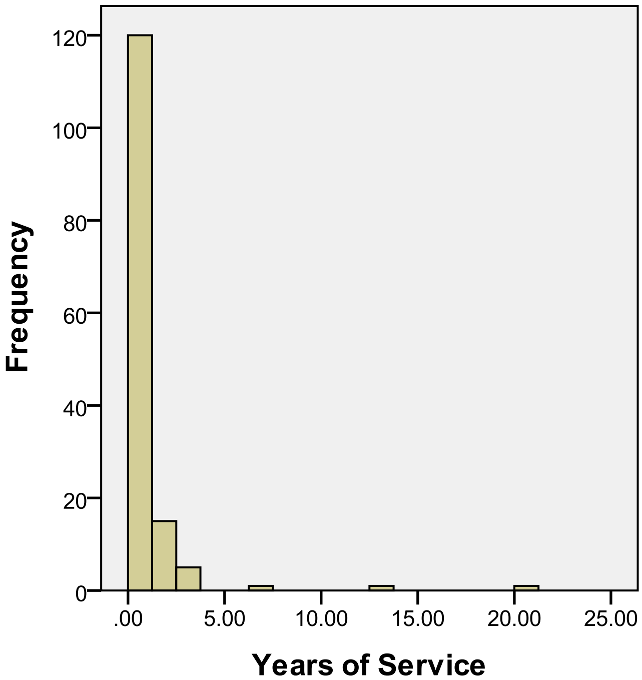


Figure 2A8. Histogram for Years of Service.

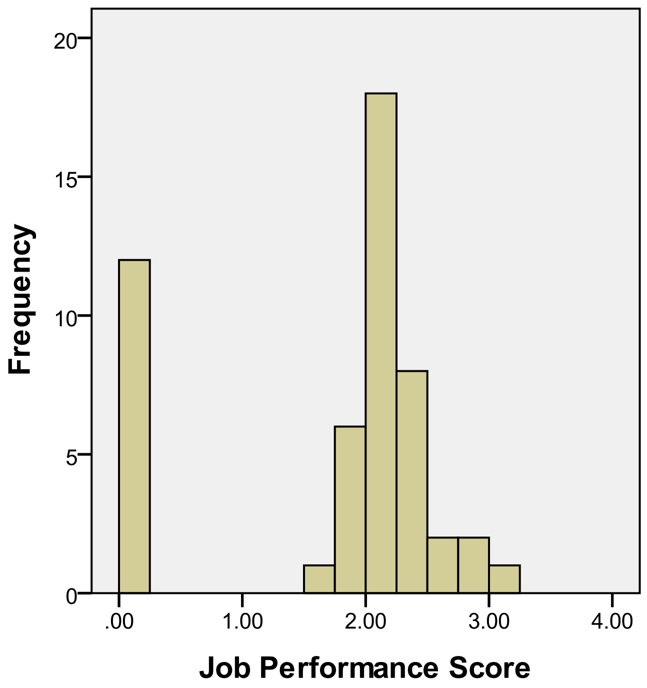


Figure 2.A9. Histogram for Job Performance scores.

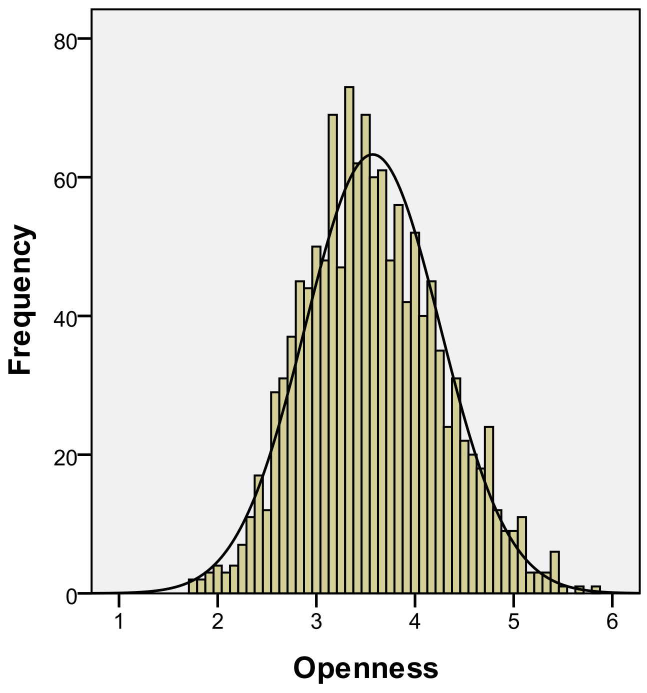


Figure 2.A10. Histogram for Openness scores.

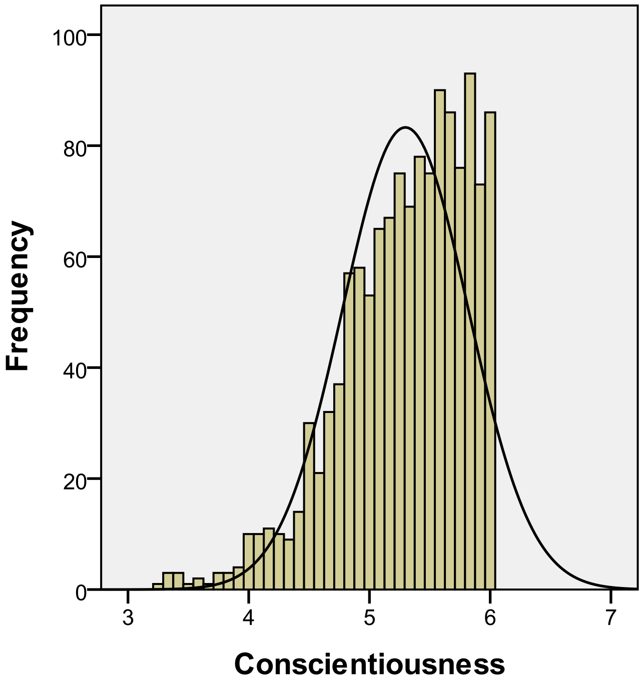


Figure 2.A11. Histogram for Conscientiousness scores.

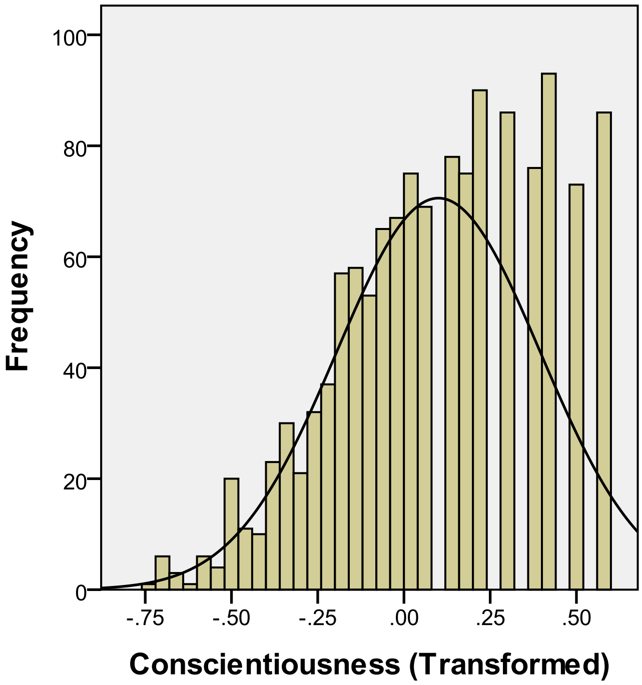


Figure 2.A12. Histogram for transformed Conscientiousness scores.

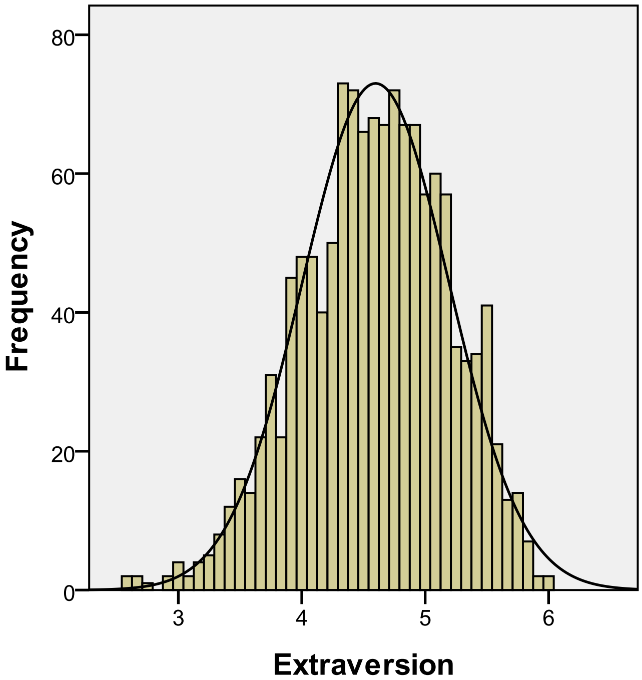


Figure 2.A13. Histogram for Extraversion scores.

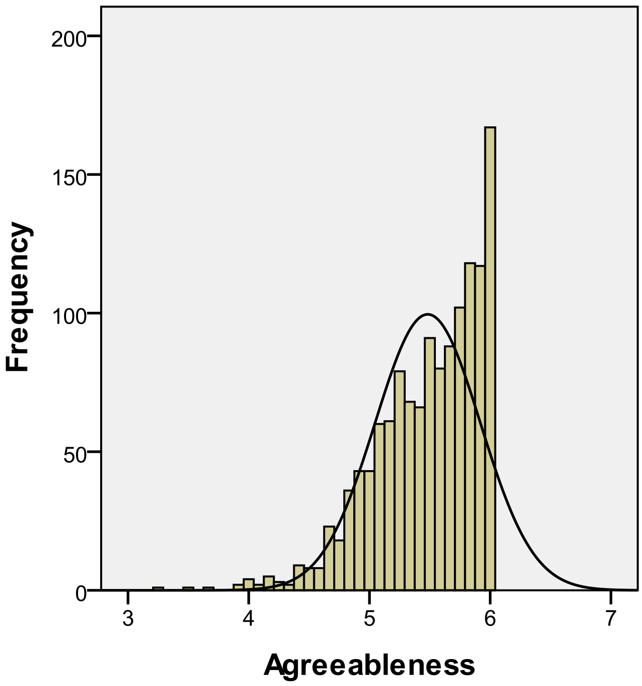


Figure 2.A14. Histogram for Agreeableness scores.

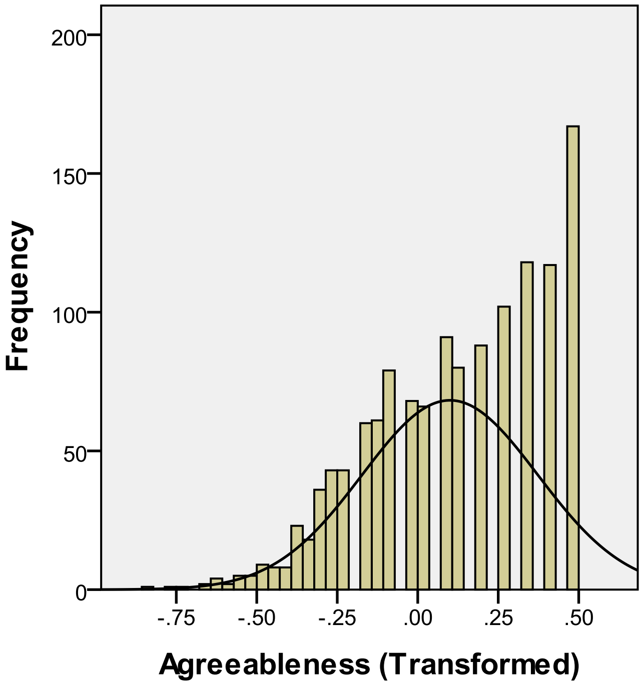


Figure 2.A15. Histogram for transformed Agreeableness scores.

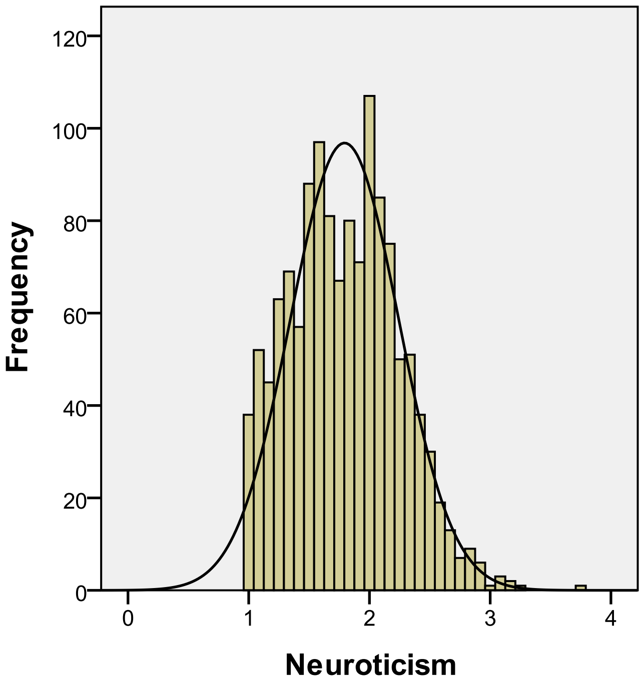


Figure 2.A16. Histogram for Neuroticism scores.

Appendix 2B

Composition of Parcels for Study 2

Parcels for Multi-Tasks Tests

5. Balls 1 – 2008, 2010, 2011, 2012
6. Balls 2 – 2013, 2014, 2015, 2016
7. Balls 3 – 2017, 2018, 2019, 2020
8. Balls 4 – 2021, 2022, 2023

Parcels for Matrices

1. Matrix 1 – 384, 376, 378
2. Matrix 2 – 387, 379, 383
3. Matrix 3 – 382, 380, 377

Parcels for Reading Comprehension

1. Read 1 – 598, 601, 615, 597, 600
2. Read 2 – 614, 604, 602, 605, 599
3. Read 3 – 595, 619, 618, 596
4. Read 4 – 603, 612, 606, 613

Parcels for Digit Sequence

1. Sequence 1 – 1971, 1771, 1821
2. Sequence 2 – 1921, 1621, 1871
3. Sequence 3 – 1471, 1671, 1571

Parcels for Word Reasoning

1. Reason 1 – 390, 403, 401, 414, 399
2. Reason 2 – 402, 415, 413, 405, 411
3. Reason 3 – 410, 398, 400, 408
4. Reason 4 – 412, 404, 396, 407

Parcels for Digit Span

1. Span 1 – 2062, 2058, 2055
2. Span 2 – 2063, 2059, 2057
3. Span 3 – 2060, 2061, 2056

Appendix 2C

Confirmatory Factor Analysis Results for Study 2

Table 2.C1

Path Coefficients for the Multi-Tasks Measurement Model

Path	B	SE	β	CR	Sig.
Multi-Tasks word order to:					
Parcel 4	1.00		.75		
Parcel 3	1.07	.04	.77	25.70	.000
Parcel 2	1.06	.04	.79	26.26	.000
Multi-Tasks placement keeping to:					
Parcel 4	1.00		.76		
Parcel 3	.56	.02	.76	24.87	.000
Parcel 2	.82	.04	.70	23.21	.000

Table 2.C2

Covariances between Constructs for the Multi-Tasks Measurement Model

Relationship	B	SE	CR	Sig.
Word order and placement keeping	.16	.01	15.68	.000

Appendix 2D

Path Coefficients for the Full Measurement Model for Study 2

Table 2.D1

Path Coefficients for the Full Measurement Model

Relationship	B	SE	β	CR	Sig.
Fluid intelligence to:					
Multi-Tasks placement keeping	1.00		.82		
Multi-Tasks word recall	.62	.04	.89	15.73	.000
Digit span	.39	.03	.64	14.60	.000
Crystallized intelligence to:					
Digit sequence	1.00		.60		
Matrices	2.01	.35	.65	5.74	.000
Reading comprehension	2.96	.50	.94	6.05	.000
Word reasoning	2.41	.39	.70	6.21	.000

Appendix 2E

OCEANIC Personality Inventory questions

The OCEANIC Personality Inventory was administered on computer. The instructions given and questionnaire are outlined below.

Instructions

The next assessment will ask you to provide information about your personal style and attitudes. Give your first impression of whether each statement describes the way you think and feel about it. Don't spend too long on deciding what your answer should be.

Select your answers by clicking on your choice using the mouse. Ask for assistance if required.

It is also important to select your answers carefully because you cannot go back and change your answer. There are NO right or wrong answers.

Please begin.

1. I am interested in all fields of science.	Never Rarely Sometimes Often Usually Always
2. I am organised.	Never Rarely Sometimes Often Usually Always
3. I like parties, where there are a lot of people.	Never Rarely Sometimes Often Usually Always
4. I am a kind person.	Never Rarely Sometimes Often Usually Always
5. When under great stress, I feel like I might break down.	Never Rarely Sometimes Often Usually Always
6. I would describe myself as a deep person.	Never Rarely Sometimes Often Usually Always
7. I feel that I am careful.	Never Rarely Sometimes Often Usually Always
8. I am talkative.	Never Rarely Sometimes Often Usually Always
9. I am considerate of the feelings of others.	Never Rarely Sometimes Often Usually Always
10. I feel jittery and tense.	Never Rarely Sometimes Often Usually Always
11. I am philosophical.	Never Rarely Sometimes Often Usually Always
12. If I start something, I work until it is finished to my satisfaction.	Never Rarely Sometimes Often Usually Always
13. I consider myself a sociable person.	Never Rarely Sometimes Often Usually Always
14. I try to be kind to everyone I know.	Never Rarely Sometimes Often Usually Always
15. I am moody.	Never Rarely Sometimes Often Usually Always
16. I have intellectual curiosity.	Never Rarely Sometimes Often Usually Always
17. I am a perfectionist.	Never Rarely Sometimes Often Usually Always

18. I am bold.	Never Rarely Sometimes Often Usually Always
19. I like to help others, even if there is nothing in it for me.	Never Rarely Sometimes Often Usually Always
20. I do not accept criticism very well.	Never Rarely Sometimes Often Usually Always
21. I think about the wonders of nature.	Never Rarely Sometimes Often Usually Always
22. I try to do more than is expected of me.	Never Rarely Sometimes Often Usually Always
23. I am a shy person.	Never Rarely Sometimes Often Usually Always
24. I consider myself a sympathetic person.	Never Rarely Sometimes Often Usually Always
25. I get very upset when I am criticized.	Never Rarely Sometimes Often Usually Always
26. I spend time in meditation and deep thought.	Never Rarely Sometimes Often Usually Always
27. I like to be consistent.	Never Rarely Sometimes Often Usually Always
28. I am bashful.	Never Rarely Sometimes Often Usually Always
29. I am considered by others to be a very friendly person.	Never Rarely Sometimes Often Usually Always
30. I worry about how things might go wrong.	Never Rarely Sometimes Often Usually Always
31. I enjoy reading poetry.	Never Rarely Sometimes Often Usually Always
32. Efficient describes me well.	Never Rarely Sometimes Often Usually Always
33. I like to strike up conversations with strangers.	Never Rarely Sometimes Often Usually Always
34. I like to be generous.	Never Rarely Sometimes Often Usually Always
35. I worry more than most people.	Never Rarely Sometimes Often Usually Always
36. I think about the origins of the universe.	Never Rarely Sometimes Often Usually Always
37. I like to have a place for everything and everything in its place.	Never Rarely Sometimes Often Usually Always
38. I take charge in group meetings.	Never Rarely Sometimes Often Usually Always
39. I have sympathy for others who are having problems.	Never Rarely Sometimes Often Usually Always
40. I am a jealous person.	Never Rarely Sometimes Often Usually Always
41. People consider me innovative.	Never Rarely Sometimes Often Usually Always
42. I like to be precise.	Never Rarely Sometimes Often Usually Always
43. At social functions, I talk to as many people as possible.	Never Rarely Sometimes Often Usually Always
44. I try to be polite, even to those who are not polite to me.	Never Rarely Sometimes Often Usually Always
45. I am an envious person.	Never Rarely Sometimes Often Usually Always
46. I spend time analysing my internal feelings.	Never Rarely Sometimes Often Usually Always
47. When I get an assignment, I do my best.	Never Rarely Sometimes Often Usually Always
48. I feel that I am withdrawn.	Never Rarely Sometimes Often Usually Always
49. I am unkind.	Never Rarely Sometimes Often Usually Always
50. I am sad and depressed.	Never Rarely Sometimes Often Usually Always
51. I am inventive.	Never Rarely Sometimes Often Usually Always
52. I like to keep all my belongings neat and orderly.	Never Rarely Sometimes Often Usually Always

53. I am quiet.	Never Rarely Sometimes Often Usually Always
54. I consider myself an unsympathetic person.	Never Rarely Sometimes Often Usually Always
55. I get upset easily.	Never Rarely Sometimes Often Usually Always
56. I enjoy visiting art museums.	Never Rarely Sometimes Often Usually Always
57. I am thorough.	Never Rarely Sometimes Often Usually Always
58. I laugh a lot.	Never Rarely Sometimes Often Usually Always
59. I am difficult to get along with.	Never Rarely Sometimes Often Usually Always
60. I consider myself a nervous person.	Never Rarely Sometimes Often Usually Always

Appendix 2F

Linear Regression results for Multi-Tasks and Personality for Study 2

Table 2.F1

Linear Regression Results for Personality Traits and Multi-task Placement Keeping (N = 1306)

Test	B	F	Sig.	Tolerance
Openness	.01	.09	.977	.88
Conscientiousness	-.02	.28	.599	.66
Extraversion	-.01	.03	.857	.80
Agreeableness	-.00	.00	.976	.67
Neuroticism	-.01	.04	.844	.74

Note. Overall model F (5, 1300) = .09, p = .995. Overall model R^2 = .000.

Table 2.F2

Linear Regression Results for Personality Traits and Multi-task Placement Keeping (N = 1306)

Test	B	F	Sig.	Tolerance
Openness	-.00	.00	.951	.88
Conscientiousness	-.04	1.44	.231	.66
Extraversion	.02	.61	.436	.80
Agreeableness	.03	.89	.346	.67
Neuroticism	.01	.12	.734	.74

Note. Overall model F (5, 1300) = .47, p = .802. Overall model R^2 = .002.

Appendix 3A

Histograms of Variables for Study 3

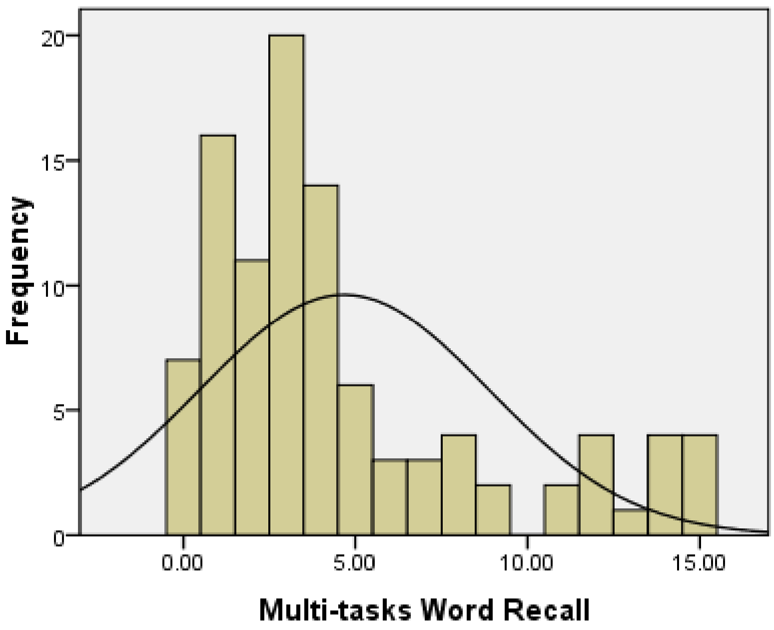


Figure 3.A1. Histogram for the Multi-Tasks Word Recall subscale.

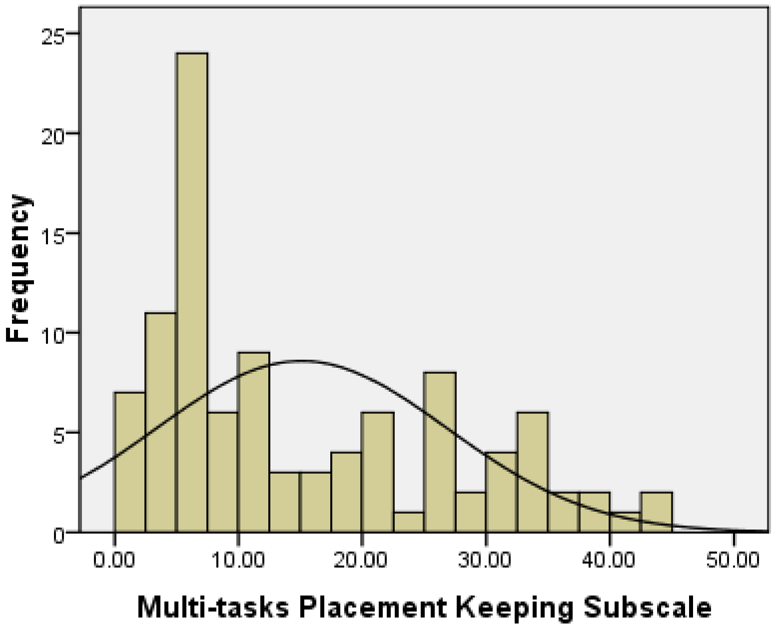


Figure 3.A2. Histogram for the Multi-Tasks Placement Keeping subscale.

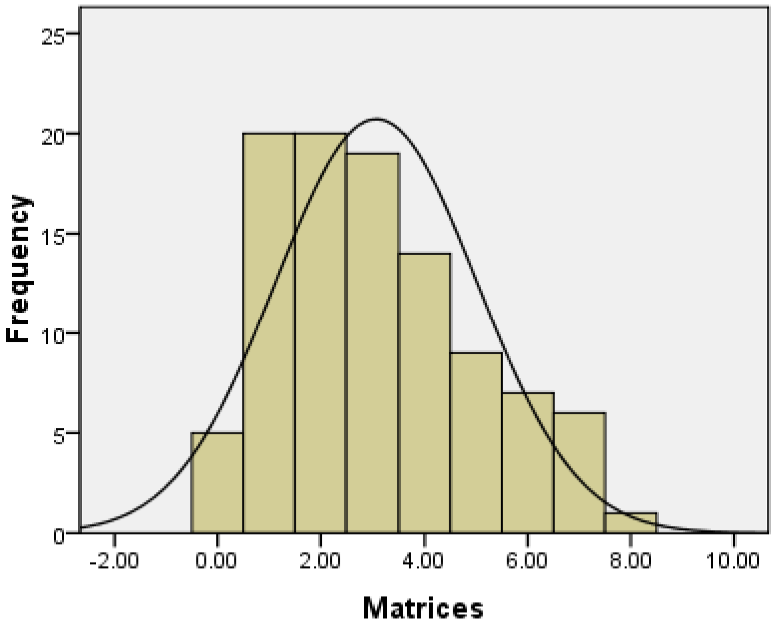


Figure 3.A3. Histogram for the Matrices scale.

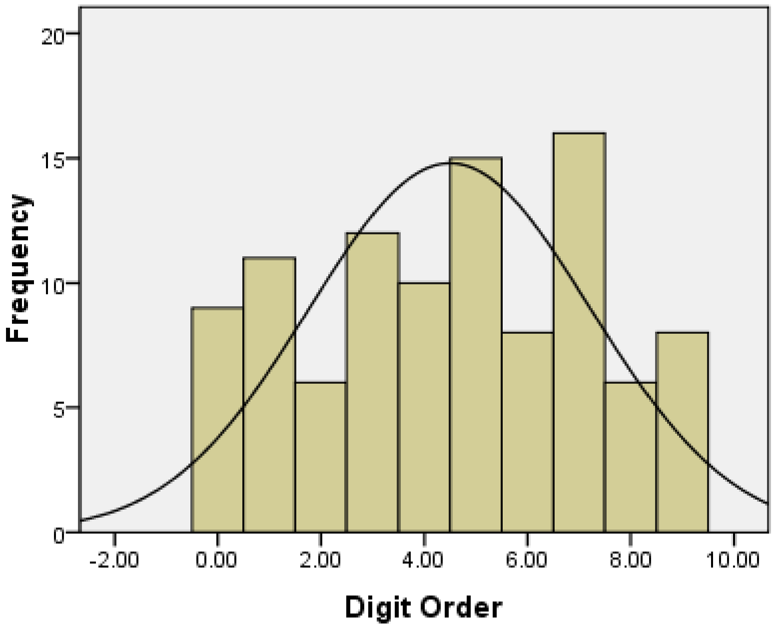


Figure 3.A4. Histogram for the Digit Order scale.

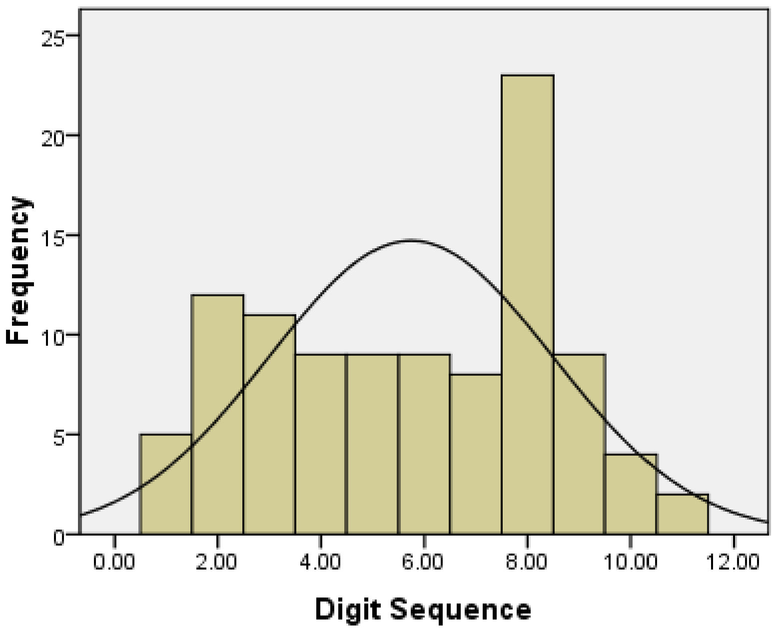


Figure 3.A5. Histogram for the Digit Sequence scale.

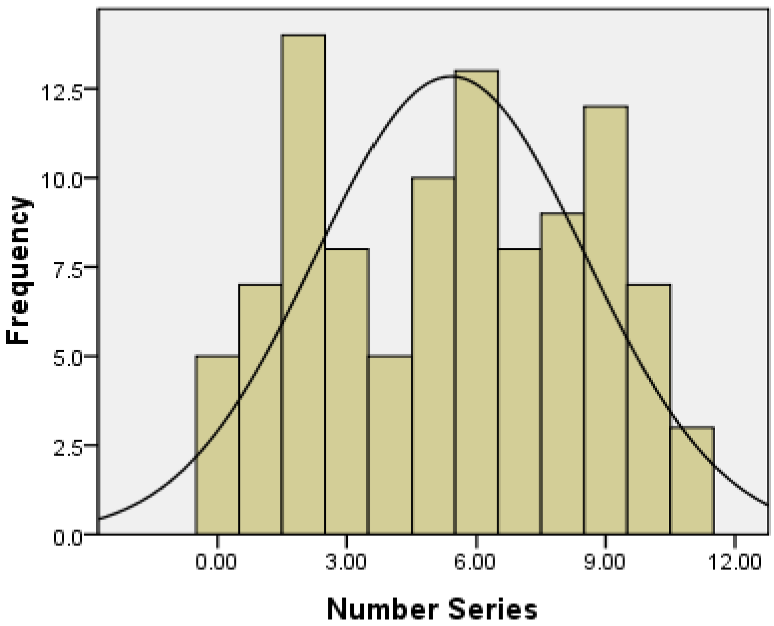


Figure 3.A6. Histogram for the Number Series scale.

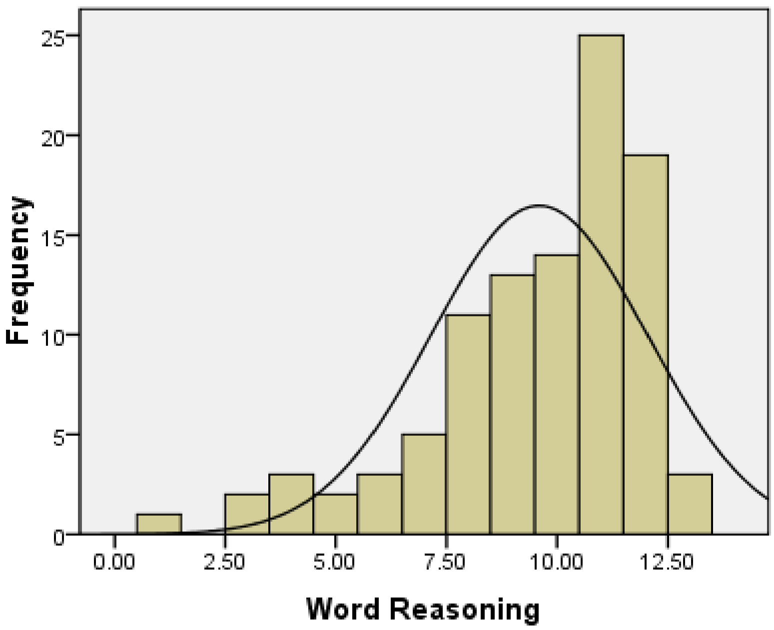


Figure 3.A7. Histogram for the Word Reasoning scale.

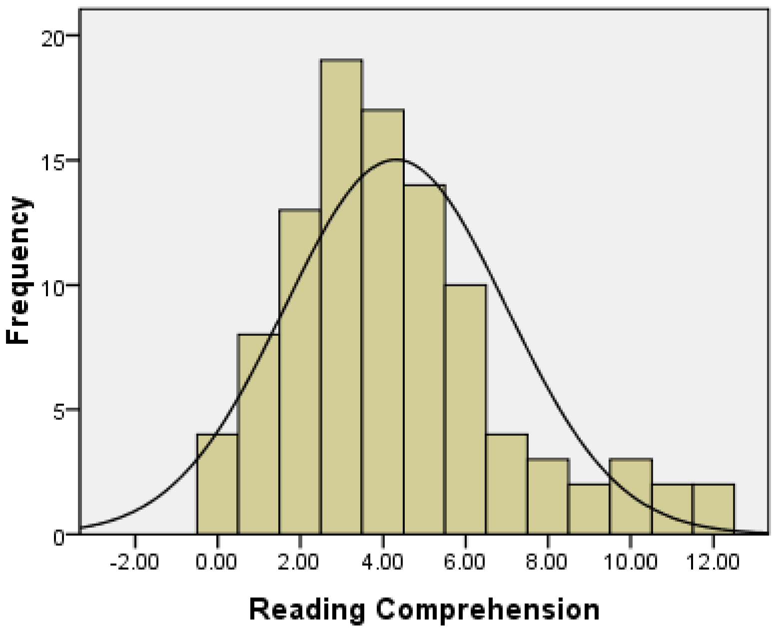


Figure 3.A8. Histogram for the Reading Comprehension scale.

Appendix 3B

Composition of Parcels for Study 3

Parcels for Multi-Tasks Tests

1. Parcel 1 – 2008, 2010, 2011, 2012
2. Parcel 2 – 2013, 2014, 2015, 2016
3. Parcel 3 – 2017, 2018, 2020, 2021
4. Parcel 4 – 2022, 2023, 2024

Parcels for Matrices

1. Parcel 1 – 378, 381, 385
2. Parcel 2 – 376, 377, 384
3. Parcel 3 – 379, 380, 382

Parcels for Digit Span

1. Parcel 1 – 2056, 2058, 2062
2. Parcel 2 – 2057, 2061, 2063
3. Parcel 3 – 2059, 2060, 2064

Parcels for Digit Sequence

1. Parcel 1 – 157, 167, 197
2. Parcel 2 – 147, 152, 192
3. Parcel 3 – 172, 182, 202
4. Parcel 4 – 162, 177, 187

Parcels for Number Series

1. Parcel 1 – 149, 174, 184, 194
2. Parcel 2 – 154, 164, 169, 189
3. Parcel 3 – 179, 199, 204

Parcels for Word Reasoning

1. Parcel 1 – 405, 408, 409
2. Parcel 2 – 396, 411, 414
3. Parcel 3 – 398, 407, 412
4. Parcel 4 – 392, 400, 404, 410

Parcels for Reading Comprehension

1. Parcel 1 – 597, 605, 614
2. Parcel 2 – 600, 601, 602
3. Parcel 3 – 606, 612, 619

Appendix 3C

Factor Loadings for the Proposed Two-Factor Multi-Tasks Measurement Model for Study 3

Table 3.C1

Unstandardised and Standardised Factor Loadings for the Two-Factor Multi-Tasks Measurement Model

Path	B	SE	β	t	Sig.
Placement keeping to:					
Keep 1	1.00		.74		
Keep 2	.96	.11	.85	8.37	.000
Keep 3	1.14	.13	.87	8.49	.000
Keep 4	.91	.11	.82	8.01	.000
Word recall to:					
Word 1	1.00		.79		
Word 2	1.24	.14	.89	9.03	.000
Word 3	1.10	.13	.79	8.24	.000
Word 4	1.60	.28	.57	5.68	.000

Appendix 3D

Path Coefficients for the Parsimonious and proposed Full Measurement Model for Study 3

Table 3.D1

Unstandardised and Standardised Factor Loadings for the Parsimonious Single-Factor Full Measurement Model

Path	B	SE	β	t	Sig.
Intelligence to:					
Placement keeping	1.00		.70		
Word recall	.60	.15	.59	4.12	.000
Digit order	1.12	.28	.60	3.97	.000
Digit sequence	.98	.26	.75	3.74	.000
Matrices	1.32	.29	.85	4.55	.000
Number series	1.52	.33	.85	4.55	.000
Reading comprehension	.98	.25	.74	3.88	.000
Word reasoning	.95	.28	.70	3.33	.000

Table 3.D2

Unstandardised and Standardised Factor Loadings for the Proposed Two-Factor Full Measurement Model

Path	B	SE	β	t	Sig.
Gf to:					
Placement keeping	1.00		.70		
Word recall	.60	.15	.58	4.10	.000
Digit order	1.12	.28	.60	3.97	.000
Digit sequence	.98	.26	.75	3.75	.000
Matrices	1.32	.29	.85	4.54	.000
Number series	1.53	.34	.85	4.55	.000
Gc to:					.000
Reading comprehension	1.00		.75		.000
Word reasoning	.97	.32	.71	3.08	

Appendix 3E

Test Administration Instructions

Test Administration Instructions for Invigilators

The main functions of an Invigilator are to ensure that an assessment commences, continues and concludes in an orderly and timely manner, and that candidates are kept under constant and effective supervision throughout an examination.

While full instructions are provided by the software program including practice examples, candidates should be reminded:

- That mobile phones, or other electronic or communication devices, are not permitted during the assessment;
- To make sure that a 'clean desk' policy applies: no pens, pencils, writing material, be brought to desks.

A log should be kept detailing any irregularities during the assessment session including:

- If there is any disturbance to candidates (e.g. noise in/outside the venue);
- If a candidate appears distressed or unwell;
- If a candidate is suspected of cheating;
- If there is an emergency in a venue.