Reasoning ability and performance:

A study of New Zealand corrections officers

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Psychology

At Massey University, Wellington, New Zealand

Bryony Jackson

2015

Abstract

The performance of frontline staff is one of the most central elements of an effective correctional system. This thesis reports findings related to the organisational psychology of the correctional environment, with special attention to person characteristics that may predict job performance of corrections officers. The empirical study investigated components of reasoning ability (abstract, verbal, numerical reasoning) on a sample of officers (N = 88) working in seven prison facilities throughout New Zealand. Overseas research repeatedly identified cognitive abilities as a predictor of job performance across a range of occupational settings, including jobs similar to corrections officers. The current study sought to examine this relationship on a New Zealand officer sample, to provide evidence for criterion-related validity of psychometric assessment of reasoning ability, with implications for use in personnel selection procedures. An analysis of internal relationships among ability components was also undertaken. Consistent with theoretical models and extant empirical findings, abstract, verbal, and numerical reasoning were found to be positively related to one another. However, the given components of reasoning ability were unrelated to job performance in the local occupational setting. Potential explanations for the findings are offered in terms of limitations in the measurement tools and processes (e.g. scope of the performance appraisal tool). It is likely that given the unique job tasks and challenges of the corrections environment, officer performance requires important characteristics outside of reasoning ability, when officers perform affect-laden tasks (e.g. understanding their own and others' emotions and emotion-driven behaviours). Further investigation of potential predictors such as emotional intelligence is warranted, and is expected to assist prediction of performance in a corrections setting.

Acknowledgments

Thank you to the Department of Corrections, including the Research and Evaluation group for granting permission to complete this thesis. Also thank you to Charlotte Stephens for the early discussions, supporting access to data, and for your original project work from which this thesis stems. Additionally I would like to acknowledge OPRA; the initial discussions with Heather Morrell and Chris Densem's support with data were kindly appreciated.

Finally thank you to my supervisor Dr Gus Habermann for all the guidance and advice provided, as well as your patience and encouragement throughout.

Contents

List of Figures

List of Tables

CHAPTER 1: COGNITIVE ABILITIES

Cognitive abilities relate to an individual's capacity to reason, plan, and solve problems (Goldstein, Princiotta, & Naglieri, 2015; Gruszka, Matthews, & Szymura, 2010). These skills or 'mental abilities' are relevant to all types of work. Within organisational psychology, cognitive abilities have received attention from a practical perspective, and studies have investigated the predictive utility of cognitive ability in selecting employees for particular job roles. There is consistent, extensive evidence that cognitive ability is associated with job performance and learning (Gatewood, Feild, & Barrick, 2015; Schmidt & Hunter, 1998).

Whereas organisational psychology has focused on cognitive ability from a practical perspective, differential and empirical psychology have discussed and debated the construct from a theoretical perspective. Researchers have attempted to define the nature and structure of both general and specific cognitive abilities. A variety of related terms are used in the literature to refer to cognitive abilities, including cognitive skills, reasoning skills, aptitudes, and more often, intelligence.

Defining cognitive ability

While distinctions have recurrently been made between the notions of intelligence, cognitive ability, and reasoning ability, a review of empirical work on cognitive ability needs to begin with considering concepts of intelligence. Theories of intelligence have evolved and developed over the last century, and these concepts have informed current theories and models of cognitive ability (Sternberg, 2000; Sternberg & Kaufman, 2011). The terms 'cognitive ability' and 'intelligence' have been used interchangeably. However, researchers presenting evidence from psychometric tools tend to distinguish between the constructs, while acknowledging their multiple links. Intelligence usually refers a general or non-specific ability to learn, reason, and problem-solve, whereas cognitive ability refers to a specific or narrower mental aptitude, such as verbal or numerical reasoning. General mental ability (GMA) is often taken to mean the same as intelligence. These disputes are methodologically rooted in the use of hypothetical constructs to account for, explain, and predict overt behaviour such as problem-solving behaviours (Cronbach & Meehl, 1955; MacCorquodale & Meehl, 1948).

Although intelligence has been discussed and debated by psychologists for more than a century, a universally accepted definition does not exist (Goldstein et al., 2015; Wilhelm & Engle, 2005). In a 1921 symposium, fourteen 'experts' provided definitions of 'intelligence'. There was little agreement about the meaning of the term, and different theorists emphasised different aspects in their definitions. For example, Terman (1921) emphasised the ability to think abstractly, form concepts, and to understand the relationships between them. Thorndike (1921) emphasised learning in his definition and defined intelligence in terms of behavioural responses to situations.

An agreement was not reached in 1921 about the meaning of intelligence, but several themes emerged, and these themes have informed later definitions of the construct. Problem-solving, learning, and reasoning were (at least implicitly) mentioned by many of the researchers. A similar symposium was held 65 years after the first. In this second symposium, themes of learning and adaptation to the environment remained in definitions. A further element arose; this was being able to understand and control oneself (Sternberg & Kaufman, 1998).

Later definitions of intelligence maintained an emphasis on an individual's interaction with the environment. Sternberg (2014) suggested that this interaction between the individual and the environment has been characteristic of most definitions, and that intelligence ultimately refers to an individual's interaction with and ability to adapt to the environment. It further involves the ability to shape the environment if the current environment is not ideal, or to select a new environment (Sternberg, 2014). Another definition was offered by Gottfredson (1997) who defined intelligence as being able to deal with cognitive complexity.

Alternative conceptualisations of intelligence have also been proposed, including the idea that multiple intelligences exist, such as intrapersonal, interpersonal, musical, and kinesthetic intelligence (Gardner, 1998, 2006), and that intelligence is a malleable construct influenced by mindset and motivation (Blackwell, Rodriguez, & Guerra-Carrillo, 2015; Dweck, 1999). The term 'intelligence' has been applied to non-traditional domains that even transcend Gardner's Multiple Intelligences, such as social and emotional intelligence (Goleman, 1995; Mayer, Salovey, & Caruso, 2004). Some researchers have argued that these domains, described as the 'hot intelligences' (Mayer, Caruso, Panter, & Salovey, 2012), may link to more meaningful life events and work outcomes than the traditional forms of intelligence. The following section discusses the evolution of models and theories of intelligence and cognitive ability.

1.1. Theoretical models and approaches

Among themes emergent in the literature on cognitive ability, perhaps the most central discussion is around the delimitation (scope) as well as componential structure of the construct, including whether it is best defined as a single, general factor, or whether there are multiple abilities (Goldstein et al., 2015; Sternberg & Kaufman, 2011; Wilhelm & Engle, 2005). Another debate is whether cognitive ability is best understood in terms of its measurement (the psychometric approach), or whether a better understanding comes through investigating the construct from a radical cognitive perspective, looking at information processing and underlying processes such as memory and retrieval. Sternberg's work on cognising in analogical reasoning items (1977), inspired by the componential approach advocated by John B. Carroll, represents the turning point when a genuine information processing approach was first reconciled with intelligence as a construct studied through correlational and factor-analytic methods.

Both the psychometric and cognitive approaches have been influential. Although the approaches vary in how they seek to understand cognitive ability, they are not necessarily mutually exclusive, and many later theories of cognitive ability have been a synthesis of the two approaches. Sternberg (2014) suggested that these different ways of studying intelligence can in fact be complementary, and the accumulation of evidence from multiple sources and research paradigms is an advantage.

1.1.1. Approaches in differential psychology: “The correlationalists”

Studies in psychometrics and differential psychology have sought to explain and define cognitive ability through its measurement (Wilhelm & Engle, 2005). In psychometric approaches, the underlying assumption is that latent individual differences in cognitive abilities manifest in test scores. It is assumed that cognitive ability is quantifiable and measurable. Early psychometric researchers analysed scores on various types of mental tests and studied the patterns of individual differences. These studies led to developments in statistical techniques, including factor analysis, to identify the latent dimensions or the ‘components’ of cognitive ability in the measured data (test scores).

A common element in most of these approaches is the assumption that evidence is to be produced by statistically analysing item-level responses. Common quantitative techniques are correlational, followed in later research by multivariate analysis (e.g. versions of factor analysis). The starting point of producing all evidence is the item-item correlation matrix, augmented in some later developments by correlating subtest scores. A tacit assumption is that research does not need to query how and why a correct or incorrect response to an item is reached by a testee.

The differential-psychological and correlational approach dominated theories of intelligence in the first half of the twentieth century, and strongly influences assessment to the present day. However, the approach dates back to the nineteenth century and Francis Galton's interest in individual differences in psychological traits. In 1869, Galton conceived of a 'general mental ability' and focused on developing objective ways to measure it (Galton, 1869, cited in Jensen, 2002). In the early twentieth century, Binet and Simon created a standardised measurement scale (Binet & Simon, 1905, cited in Carson, 2014). The Binet-Simon scale consisted of 30 brief cognitive tests designed to assess different skills, including reasoning, memory, and semantic judgements. It served as a prototype for other later revisions, including the Stanford-Binet scale (cf. Boake, 2002). Binet and Simon described intelligence as using judgement, adapting to the environment, directing one's efforts, and exercising self-criticism (Binet & Simon, 1916, cited in Sternberg & Kaufman, 1998). These aspects of intelligence continued into later definitions of the construct.

The general or g factor of intelligence

Most models describe intelligence as comprising of one or more components, which can also be described as facets, dimensions, or factors. The exact number of components is debated. The most pervasive theory in the psychometric literature is that a single or 'general' factor (*g*) underpins individual differences in cognitive ability measures. The theory of a general factor of intelligence was first proposed by Spearman (1904) who described what he provisionally called 'general intelligence'. Using correlational techniques, Spearman identified that individuals who did well in one form of school test, tended to do well in other types of test. To explain this phenomenon, Spearman (1904) suggested that all tests measure, at least in part, a general or fundamental type of intelligence.

Spearman's finding that test scores positively correlated led to further investigations into the structure of intelligence. Studies used factor analysis and consistently identified a main factor accounting for a large amount of the variance in test scores (Spearman, 1927). This phenomenon was termed the 'positive manifold'. Spearman proposed a two-factor theory of intelligence in 1927, and argued that intelligence tests measure two types of factors: the first factor was the general factor and common to all tests, and the second type was a unique factor specific to the test.

The finding that scores on different ability tests tend to correlate (or the 'positive manifold') has been identified by later researchers (cf. Jensen, 1986; Ree & Earles, 1991). The phenomenon is often explained as evidence of a general factor of intelligence (Reeve & Hakel, 2002). With factor analysis the existence of a general factor in test scores on different types of has been supported many times, and it may be one of the most replicated psychological research findings (Sternberg, 2014). Researchers have also sought to investigate whether the general factor exists cross-culturally and to rule out the possibility that it may be an artefact of factor analysis.

Some questions remain about the construct of *g*. Sternberg (2014) acknowledged that *g* existed, but argued that the more important question is how general is the *g* factor. This question referred to the relationship between *g* and specific cognitive abilities, and the issue of whether *g* accounts for most of intelligence, or only a small amount with specific abilities making up a greater proportion. Schmitt (2014) discussed how the magnitude of *g* was still under investigation, along with the degree to which measures of specific ability add in terms of predicting job performance (i.e. the incremental validity of specific abilities). Given that *g* appears to underpin the ability to learn and process information, an ability that is relevant across all job roles and settings, it is likely that *g* will remain important as a construct of interest in organisational psychology.

Specific abilities and the development of hierarchical models

Some researchers have preferred to focus on specific cognitive abilities. Thurstone (1938) questioned Spearman's concept of a general factor and investigated the concept of 'reasoning', particularly the notion that separate forms of reasoning may exist. Using factor analysis, Thurstone argued in favour of seven primary categories of ability: numerical, reasoning, spatial, perceptual, memory, verbal fluency, and verbal comprehension. Thurstone's 1938 model was an early example of a model of cognitive abilities. This model was influential in the development of later models of intelligence, particularly hierarchical models that recognised that as well as the general form, there are also specific types of cognitive ability.

Hierarchical approaches to cognitive ability propose that more general cognitive abilities are higher up in the hierarchy of abilities, and the more specific abilities are lower in the hierarchy (Sternberg & Kaufman, 1998). Carroll's (1993) hierarchical model of intelligence, the 'three-strata model', was based on the factor analysis of more than 460 data sets obtained from 1927 to 1987. Stratum 1 included multiple, narrow, specific abilities, such as spelling ability and reasoning speed. Stratum 2 covered broad abilities, including memory, and Stratum 3 was a single, general intelligence (similar to Spearman's *g* factor).

Spearman and Kaufman (1998) argued that Carroll's 1993 model did not break new ground, as it did not propose any abilities that had not been previously been suggested. However, as the model integrated a large amount of factor-analytic research, it had empirical support and a degree of authority over alternative conceptualisations. Although later models have proposed alternative structures, Carroll's hierarchy continues to provide one of the most influential taxonomies for the structure of cognitive ability (Ones, Dilchert, & Viswesvaran, 2012). Hierarchical models of cognitive ability remain popular in the literature (Schweizer, Troche, & Rammsayer, 2011).

Crystallised and fluid intelligence

Intelligence can also be described as crystallised and fluid. Fluid intelligence (*Gf*) refers to the dimension of intelligence that involves solving abstract problems, thinking logically, and identifying patterns and relationships. This form of intelligence is independent of acquired knowledge, experience, and education. In contrast, crystallised intelligence (*Gc*) is the ability to use knowledge that comes from past experiences and education. The *Gf-Gc* theory of intelligence is now widely applied in psychological research and practice (Beauducel, Brocke, & Liepmann, 2001).

The concepts of fluid and crystallised intelligence were proposed by Cattell in the mid-twentieth century and later refined through a number of studies using factor analysis of test scores (cf. Cattell, 1963; Horn & Cattell, 1966). Cattell disagreed that general intelligence was a unitary construct and proposed that general intelligence could be further categorised as fluid and crystallised (Horn & Cattell, 1966). Another way of viewing fluid and crystallised intelligence was in terms of investable (fluid) intelligence, and invested (crystallised) intelligence. Cattell's Investment Theory was that general cognitive capacity (fluid intelligence) is invested into particular areas during development, and relative differences in an individual's specific abilities are a result of these differential investments. Relative investments into different skills are driven by educational experiences, preferences, interests, and other personal characteristics (Ones, Dilchert, Viswesvaran, & Salgado, 2010). This theory can help explain the relationship between general and specific ability, as well as explaining how specific abilities might develop.

One approach to understanding fluid and crystallised intelligence is to study the constructs in relation to age. Horn (1968) built upon Cattell's theory of crystallised and fluid intelligence, focusing on how these abilities developed over the life-span. Crystallised intelligence was viewed as largely a function of experience, and was therefore expected to increase through the life span. For fluid intelligence, the general theory was that it increased throughout childhood and early adulthood, and then declined eventually. The development of neural and physiological structures was accompanied by an increase in fluid intelligence.

Although the development of fluid intelligence was dependent on learning, learning itself was dependent on developing of the physiological structures that supported it (Horn & Cattell, 1967). Horn and Cattell suggested that the decline in neural function that occurs throughout adulthood had a greater impact on fluid intelligence than crystallised intelligence.

One of the debates surrounding fluid and crystallised intelligence is around structure, for example, whether they are higher- or lower-order factors, or whether broader or narrower conceptualisations offer a better understanding. For instance, in Carroll's three-stratum model, crystallised and fluid intelligence were in the second stratum, with the broad abilities. However, other conceptualisations have equated fluid ability with general mental ability or *g* (Gustafsson, 1984; Valentin Kvist & Gustafsson, 2008). For example, Valentin Kvist and Gustafsson (2008) found that the constructs of *g* and *Gf* were almost perfectly correlated in groups that were homogenous in terms of background knowledge and experiences (Swedish non-immigrants, European immigrants, and non-European immigrants), but the correlation was .83 when the subgroups were treated as one group. Findings such as this raise the question of whether it is possible to measure fluid ability separately from general mental ability.

Alternative interpretations have been proposed for both fluid and crystallised intelligence, suggesting that these forms of intelligence may include specific facets, or may themselves be facets of broader dimensions. Additionally, these intelligences may involve overlapping and faceted structures (Beauducel et al., 2001). For example, potential facets of *Gf-Gc* would include forms of reasoning (for fluid intelligence) and different types of knowledge and language aptitudes (for crystallised intelligence). When taking a broader view, fluid intelligence comes to include both memory and perceptual speed, and crystallised intelligence comes to include knowledge and knowledge fluency (Beauducel et al., 2001).

The degree of overlap and relationship between *Gf* and *Gc* is also debated. *Gf* and *Gc* are correlated (Lubinski, 2004) and can therefore be thought of as related forms of abilities. Other researchers view them as more distinct. Nisbett et al. (2012) explored the research to date and summarised evidence to support the theory that fluid and crystallised intelligence are different forms of abilities. This evidence included findings from brain research. At the biological level, fluid intelligence appeared to be substantially mediated by the prefrontal cortex, whereas crystallised intelligence was not. This was consistent with the finding that fluid intelligence declines with age more so than crystallised intelligence, as the prefrontal cortex also declines more rapidly than other parts of the cortex. Nisbett et al. described other findings to support the proposal that fluid and crystallised intelligence are quite separate, including that training in executive brain functions, including working memory, can have a great effect on fluid intelligence but not on crystallised intelligence.

1.1.2. Current cognitive approaches

Psychometric approaches were popular in explaining intelligence in the first half of the twentieth century. These approaches were also criticised on a number of grounds. One criticism was the inability of correlational approaches – prominent in psychometric work – to identify the processes that contributed to task performance, and the recognition that understanding process as well as structure is important in understanding the nature of intelligence (Sternberg, 1983). Arguments such as these led to increased interest in information processing and cognitive theories of intelligence (Sternberg & Pretz, 2005).

Cognitive theories of intelligence focus on the processes underlying individual differences in task performance. As well as identifying the processes, cognitive theories also investigate the speed and execution of the processes. Whereas psychometric approaches are concerned with the number of factors in intelligence, cognitive approaches are concerned with 'levels of processing'. The debate in the cognitive approach is often around how many levels are involved in processing, and most theories and models within this approach propose that processing occurs at a number of levels.

Speed of processing

In an information processing perspective, the focus is on individual differences in processes and operations per se, rather than information content (Gruszka et al., 2010). Therefore, higher scores on intelligence tests reflect greater quality of processes, rather than simply knowing content. Information processes include concepts such as working memory and reaction times.

Jensen (1992) did not dispute the existence of a general factor of intelligence, but argued that it was best understood in terms of individual differences in information processing. Jensen pointed out that *g* correlated with dimensions outside of psychometrics and factor analysis. One of these correlates was reaction time. Interestingly, reaction time showed greater correlation with non-speeded intelligence tests than speeded intelligence tests, suggesting that the natural speed at which an individual answered correctly was associated with correct response. Individual differences in inspection time (time taken to make a simple visual or auditory discrimination) were also correlated with *g*.

Jensen also discussed correlates of *g* in terms of physiology and brain structures. The glucose metabolised by the brain while undertaking a highly *g* loaded test of abstract reasoning was negatively correlated with the individual's test scores. The speed of neural transmission in the neural tract, from the retina to the visual cortex, was also significantly correlated with scores on Raven's Progressive Matrices. This last finding was of particular theoretical importance, because the visual tract is not involved in higher mental functions. Findings such as these raise the possibility that individual differences in cognitive abilities and performance may at least in part be accounted for by individual differences in information processing and physiological processes.

Planning, Attention, Simultaneous, Successive Theory

Based on findings from neuropsychology, Naglieri and Das (1990) argued that intelligence was better understood as the outcome of interrelated cognitive processes (planning, attention, simultaneous processing, and successive processing) through three functional units. The first functional unit was associated with cortical arousal, attention, and planning; the second would code information simultaneously or successively; and the third was associated with self-monitoring and developing cognitive strategies. The different processes involved would depend on the task, for example, mathematics would involve different contributions from the cognitive processes than reading.

The theory, named Planning, Attention, Simultaneous, Successive (PASS), challenged psychometric theories, particularly the notion of *g*. The researchers argued in favour of using the term *cognitive process* as a more modern term for *ability* (Naglieri, Das, & Goldstein, 2012). The PASS model and the associated Cognitive Assessment System (CAS) have also been challenged, particularly based on findings from factor analysis. Using confirmatory factor analysis, Kranzler, Keith, and Flanagan (2000) found that the PASS model did not provide a better fit to the data than other models of cognitive ability, and the planning and attention factors were indistinguishable.

Development of a componential theory

The prominent psychometrician and psycholinguist Carroll pioneered a componential approach to explain the different ways testees tackle cognitive problems on ability and intelligence tests. Based on these proposals, Sternberg (1983) outlined a radically 'componential' model of human intelligence that combined elements of both psychometric and information processing approaches. The model proposed that intelligence is understood in terms of the functions of its components and the types of interactions between the components. The components in the model were processes, and each component process had a type of function.

Triarchic Theory of Intelligence

Sternberg's (1984a, 1985) Triarchic Theory intensified the use of findings from cognitive science, while broadening the scope of phenomena covered by a theory of intelligence. The Triarchic Theory is a meta-theory consisting of three separate theories that do not correspond to component constructs. Sternberg (1984b) believed that intelligence as operationalised by mainstream psychometric tests was narrow, and that expanding the domain of intelligence would improve predictions.

The first sub-theory, Componential, related to cognitive architecture and process, with references to psychophysiology and cognitive neuroscience. It included a taxonomy of information processing components (metacomponents, performance components, and knowledge acquisition components). At the top level, metacomponents ensure executive planning, resource allocation, and decision-making. Ten types of metacomponents were proposed and included higher-order processes such as recognising that a problem existed, selecting strategies to solve the task, and acting on feedback. The second type of components were the lower-order performance components, controlled by the metacomponents and executing narrower tasks. The last type of component was the knowledge-acquisition component, which integrated a theory of human learning with the intelligence model.

Sternberg’s view of intelligence was unique at the time, due to its focus on understanding the cognitive processes underlying performance on tasks. The model proposed that task performance was achieved through interacting processes (the components) at different levels. Depending on the task, the components involved would vary. For example, metacomponents were broadly applicable, whereas other components were more relevant to particular kinds of tasks. The model also attended to information-processing constraints.

The second sub-theory, Experiential, related to a continuum of entrenchment of a task. Entrenchment and non-entrenchment describe relative levels of experience an individual has with a type of task (finding it novel, or being able to mobilise an automated sequence). The experiential theory included a thorough analysis of the nature of insight in problem-solving.

The third sub-theory, Contextual, brought the theoretical basis of environment-specific cognition into theorising on intelligence. It related problem-solving to the external environment, utilising the classic Banduran model of an individual first adapting to, then adjusting his or her environment, and finally choosing a different environment proactively.

The components of triarchic intelligence were re-labelled (1) componential or analytic intelligence, (2) creative intelligence, and (3) practical intelligence (the worth of this simplification is disputed by Sternberg himself as well as his critics). Traditional 'general intelligence' was considered part of analytic intelligence in the model.

An important breakthrough of Sternberg's theory was the strengthened focus on context and environment in determining adaptive behaviour. Influenced by Piagetian theory, context and adaptation were highlighted in the third sub-theory of the triarchic model. Sternberg emphasised the adaptive characteristics of a person, or what might be called 'successful intelligence' in a specified field of praxis (1996, cited in Sternberg & Kaufman, 1998). Successful intelligence can be described as the ability to adapt to, shape, and select environments that meet one's own goals, as well as those of society and one's culture (Sternberg, 2012, 2014). Sternberg (2004) raised the possibility that practical skills, although just as necessary for adaptation to the environment, may go undetected in assessment via traditional standardised tests. However, these skills may be particularly relevant for predicting adaptation to everyday environments.

1.1.3. A few alternative approaches

Intelligence as process, personality, interests, and knowledge theory

Ackerman's (1996) intelligence as process, personality, interests, and knowledge (PPIK) theory incorporated *Gf*- and *Gc*-types of abilities, but extended the theory in two ways. Firstly, it included the interaction of personality and interests with intelligence facets, and secondly it included a developmental aspect to explain how the abilities developed over time. According to the theory, general intelligence is comprised of two main factors: intelligence as process (*Gp*) and intelligence as knowledge (*Gk*). Although relating to *Gf*-type abilities, *Gp* was linked to information processing and cognitive processes. *Gk* was conceptually similar to *Gc,* but broadened to include a wider range of knowledge domains.

The developmental component of Ackerman's theory was informed by Cattell's Investment Theory. In PPIK theory, individuals have varying levels of *Gp*, and through the interaction of *Gp* with personality and interests, individuals devote more or less effort to acquiring knowledge in particular domains. The knowledge domains vary during different stages of life and include academic (e.g. mathematical) and occupational domains.

The PPIK theory was also one of the first theories to broaden the intelligence construct to include the broader domain of knowledge and skills, and to integrate personality and cognitive ability. It can be considered example of a 'meta-theory' that accounts for the interplay between cognitive and non-cognitive factors. Ackerman's model may help explain how common psychological characteristics can give rise to unique cognitive ability profiles (Lievens & Reeve, 2012). The model has received empirical backing through findings that have supported the predicted trajectory of *Gp* and *Gk* by age, as well as the correlations between selected personality traits and cognitive investment in particular domains (Ackerman & Beier, 2003).

Cattell-Horn-Carroll theory of cognitive abilities

The Cattell-Horn-Carroll (CHC) theory synthesised two main models of cognitive ability: fluid and crystallised intelligence (cf. Cattell, 1963; Horn & Cattell, 1966) and the three-stratum theory (Carroll, 1993). It was developed with the intent of providing a taxonomy that reflected previous research on cognitive abilities and to provide a common framework for researchers (Schneider & McGrew, 2012). The Cattell-Horn-Carroll theory was developed using factor analysis and is arguably the most empirically supported psychological theory of intelligence.

Similar to other hierarchical models of intelligence, the Cattell-Horn-Carroll model has general intelligence (*g*) at the apex. The model incorporates *g* as the broadest of the cognitive ability constructs; however, recognising that *g i*s debated, some researchers have suggested ignoring theoretical *g* if it has no merit in a particular applied setting (Schneider & McGrew, 2012). The next stratum in the model was termed 'domain-free general capacities' and consisted of broad or general abilities. Schneider and McGrew (2012) proposed that the broad abilities include fluid reasoning (*Gf*), short-term memory (*Gsm*), long-term storage and retrieval (*Glr*), processing speed (*Gs*), reaction and decision speed (*Gt*), psychomotor speed (*Gps*), domain-specific knowledge (*Gkn*), reading and writing (*Grw*), quantitative knowledge (*Gq*), visual processing (*Gv*), auditory processing (*Ga*), olfactory abilities (*Go*), tactile abilities (*Gh*), kinaesthetic abilities (*Gk*), and psychomotor abilities (*Gp*).

Each of the general capacities was theorised to contain a range of narrower abilities. For example, the general capacity of fluid reasoning (*Gf*) could be described as the ability to solve novel, 'on-the-spot' problems (Schneider & McGrew, 2012). *Gf* could be broken down into narrower abilities, including induction, general sequential reasoning, and quantitative reasoning. The general capacity of short-term memory (*Gsm*) could be further refined in terms of memory span and working memory capacity. Unresolved questions remain around some of the domains, particularly in the narrower abilities, and researchers have acknowledged that CHC theory will undergo continuous refinements and updates (Schneider & McGrew, 2012).

Recent research has called for a renewed interest in the multidimensional nature of intelligence, including research into specific cognitive abilities (Schneider & Newman, 2015). This viewpoint argues that multidimensional and hierarchical models provide a better empirical fit than models that view intelligence as a unidimensional construct. Additionally, alternative explanations (other than *g*) have been proposed to account for the observed positive manifold phenomenon. One of these theories is that the that positive manifold arises from beneficial relationships between cognitive processes (van der Maas et al., 2006). There is also recognition that multiple processes and factors may produce the positive manifold (Horn & Blankson, 2012).

Multiple Intelligences model

Some psychologists have argued that it is more meaningful to conceptualise intelligence as multiple abilities, rather than a single factor. Gardner's theory of multiple intelligences shifted away from the *g* factor theory, and towards specific and multiple facets of intelligence. Gardner (1998) suggested individuals are best viewed as having multiple kinds of intelligences, instead of the single form assessed by tests of general mental ability. Gardner did not necessarily dispute the existence of *g*, but argued that acknowledging multiple intelligences could potentially provide for better educational outcomes and have implications for schooling (Gardner & Hatch, 1989).

Gardner proposed that there are least eight to nine types of intelligence: linguistic, logical-mathematical, musical, spatial, bodily-kinaesthetic, personal (interpersonal and intrapersonal), naturalist, and existential (Gardner, 1998). The theory of multiple intelligences stated that individuals could have more or less of each type of intelligence, and therefore a multitude of intelligence profiles could exist. This viewpoint was a significant departure from the *g* factor model. Gardner also criticised the reliance on tests to measure intelligence.

Gardner's theory of multiple intelligences has been criticised by a number of researchers, mostly on the grounds of lacking in empirical support. A study was conducted to investigate Gardner's theory of multiple intelligence in a sample of 200 adults (Visser, Ashton, & Vernon, 2006a). Tests to assess each of the eight hypothesised intelligences were administered, along with the Wonderlic Personnel Questionnaire (general intelligence). Using factor analysis, the researchers found a large *g* factor. The factor loadings were particularly strong for the intelligences with a more cognitive orientation (linguistic, logical/mathematical, spatial, naturalistic, interpersonal), and lower loadings were found for other ability tests involving sensory or motor skill (bodily-kinaesthetic intelligence).

Visser et al. (2006a) argued that the high correlations between the different intelligence tests were due to their ‘reasoning’ component, therefore pointing towards the importance of the *g* factor. The authors also argued that some of the intelligences proposed by Gardner had weak psychometric support and that they were better termed ‘talents’ as opposed to intelligences (Visser et al., 2006a; Visser, Ashton, & Vernon, 2006b).

Social and emotional intelligence

Gardner’s model of multiple intelligences included the idea that intrapersonal and interpersonal skills could be forms of intelligence. The idea of social intelligence had been earlier touched upon by earlier psychologists. Thorndike wrote about a ‘social intelligence’ in the 1920s (Stein & Deonarine, 2015). Wechsler in the 1940s had urged for non-cognitive forms of intelligence, the ‘affective’ and ‘conative’ abilities, to be assessed to provide a more complete view of intelligence, although the ideas were not paid much attention at the time (Stein & Deonarine, 2015).

The theory of emotional intelligence was a significant departure from the traditional models of intelligence because the theory was entirely focused on a non-cognitive domain. Salovey and Mayer (1990) described emotional intelligence as the ability to appraise emotions both in one’s self and others and to express and regulate emotions. The scope of emotional intelligence also came to include the use of emotional information in problem-solving (Mayer & Salovey, 1993).

Rejecting suggestions that emotion and intelligence were distinct, the authors argued that emotional intelligence met the ‘criteria’ to be termed a form of ‘intelligence’ (Mayer, Caruso, & Salovey, 1999; Mayer, Salovey, Caruso, & Sitarenios, 2001). These included the findings that the construct correlated in certain ways with other forms of intelligence (e.g. that emotional intelligence was partially distinct from verbal ability), that it could be operationalised as a series of ability tests, and that emotional intelligence abilities developed with age (Mayer et al., 1999). It was also argued that correct answers existed to tests of emotional intelligence and that measures of the construct were reliable (Mayer et al., 2001).

Alternative models of emotional intelligence have aligned the construct with personality traits and dispositions rather than ability. Petrides and Furnham (2001) described ‘trait emotional intelligence’ (or trait EI). Through two studies, the authors found a distinct emotional intelligence factor in data collected using the Eysenck Personality Profiler and the NEO PI-R, and a truncated emotional intelligence factor within the Five Factor Model. The authors argued that trait EI was a composite construct in the lower-stratum of personality taxonomies (Petrides & Furnham, 2001). Other models of emotional intelligence have combined ability and trait approaches to the construct (cf. Goleman, 1995). Although the evidence is mixed regarding the nature of the construct, emotional intelligence remains a popular topic in the psychology literature, and it has been investigated in relation to a variety of different outcomes including leadership and performance at work.

1.1.4. Intelligence and culture

That intelligence is understood differently in diverse cultures is well recognised by psychologists. For example, Sternberg (2004) stated that intelligence could not be fully or meaningfully understood outside its cultural context. It has been suggested that intelligence has characteristics of a social construction or culture-dependent notion (Sternberg, 2014). Accordingly, assessments of cognitive ability that have been developed and validated in one culture may not be valid to the same degree in another culture. This is an important consideration, given that most of the literature around cognitive ability, as well as most assessments, have been developed by Western researchers. Moreover, everyday or informal understandings of 'intelligence' are likely to vary from academic descriptions and the forms of intelligence measured by assessments.

Even the act of taking an ability test can be interpreted differently across cultures. For instance, collaboration is usual and expected for Mayan children rather than answering questions individually (Greenfield, 1997). This may reflect different conceptualisations of the self and of abilities in individualistic and collectivist cultures. Nisbett (2003, cited in Sternberg, 2004) found that Asian cultures tended to be more dialectical in their thinking, relative to Western cultures who were more linear. People from different cultures may construct concepts in different ways. The implication is that what appears to be differences in test scores, may actually reflect differences in cultural properties. For example, sorting taxonomically may seem logical in one culture, but not in another culture where they would sort functionally.

Yang and Sternberg (1997) examined Chinese conceptualisations of intelligence and found five underlying factors: (a) a general factor, (b) interpersonal intelligence, (c) intrapersonal intelligence, (d) intellectual self-assertion, and (e) intellectual self-effacement. In a study by Chen (1994) three factors were found: nonverbal reasoning ability, verbal reasoning ability, and rote memory. Interestingly, these Chinese conceptualisations related to peoples' conceptions of intelligence, otherwise known as implicit theories of intelligence. A further example that highlights cross-cultural variation in views of intelligence is the emphasis on speed of processing in Western cultures, whereas non-Western cultures place greater importance on depth of processing. In some cultures, there may even be wariness associated with the quality of work undertaken with speed (Sternberg & Kaufman, 1998).

1.1.5. Cognitive ability in organisational psychology

Defining cognitive ability at work

The organisational psychology literature on cognitive ability is largely focused on prediction (e.g. predicting job performance through cognitive ability tests in selection), and methodological issues around measurement. There is relatively less emphasis within organisational psychology on issues relating to proposed constructs of cognitive ability.

In the 1950s, the industrial psychologist Fleishman began analysing abilities in relation to work and started developing a set of dimensions or taxonomy of abilities that could be applied to a large variety of work settings. Fleishman found the distinction between 'cognitive' and 'non-cognitive' tasks in the literature at the time was too broad and thought that further dimensions were needed (Fleishman, 1967, 1975). The taxonomy, which was expanded over time, came to include cognitive, physical, psychomotor, sensory-perceptual, and social-interactive abilities. The cognitive abilities included oral comprehension, fluency of ideas, memory, numerical ability, along with inductive and deductive reasoning. Fleishman (1972, 1975) distinguished ability from skill and described ability as a general capacity rated in performance in a variety of tasks, whereas skill was proficiency in a task. Fleishman's contribution in the cognitive ability field was in developing a way of describing abilities at work and evaluating their use in applied (work) settings.

Intellectual competence

A more recent approach in organisational psychology has been to look at competency models that focus on the links between networks of constructs. For example, it has been found that measures of personality types and traits can predict cognitive ability scores in business settings (Furnham, Dissou, Sloan, & Chamorro-Premuzic, 2007). The implication is that both cognitive ability and personality will provide a better understanding of competence and potential at work.

Chamorro-Premuzic and Furnham's (2004; 2006) model of intellectual competence is an example of a model that links a network of constructs, including the effects on work outcomes. Specifically, the model links intellectual competence with intelligence (psychometric), personality, work performance, academic performance, and self-assessments of intelligence. The advantage of models such as this over psychometric approaches is that it links the construct to real-world work outcomes, as well as provides a more updated view of intelligence than that provided through psychometric testing.

Criticisms and future directions

It has been argued that the focus in organisational psychology on cognitive ability as a predictor has been at the expense of investigating the nature and structure of the construct in depth, including how to measure it and its implications for workplace behaviour (Scherbaum, Goldstein, Yusko, Ryan, & Hanges, 2012). Despite there being debate and discussion around the psychometric approach in the wider psychological domain (e.g. in neuropsychology and educational psychology), organisational psychology has remained focused on psychometric conceptualisations of intelligence (Scherbaum et al., 2012).

There are several potential reasons why most of the organisational psychology literature on cognitive ability has a constrained scope. One reason is that the relationship between general mental ability and job performance is established (cf. Schmidt & Hunter, 1998). Researchers may see little point in investigating other angles and components of ability (Scherbaum et al., 2012). A real-life reason for a focus on prediction and validation is that it is highly important in the legal environment (e.g. that of North America) to prove that the use of psychometric tests in personnel selection is justified. There is greater scrutiny to ensure that tests are valid, for example, to prevent discrimination in hiring and lawsuits (Scherbaum et al., 2012).

Scherbaum et al. (2012) acknowledged that definitions of intelligence vary widely, but that different areas tend to emphasise different aspects in their definitions. It follows that a unified definition of intelligence is required for organisational psychology. That is, an understanding of intelligence as it pertains to workplace settings. Lievens and Reeve (2012) also suggested that organisational psychology needs to adopt a broader view of intelligence, and encouraged organisational psychologists to recognise that intelligence was not a unitary construct, but instead refers to a network of constructs including cognitive abilities and skills.

Lievens and Reeve (2012) explained that there are two types of intelligence that can be distinguished: (a) the ability to learn and solve problems (i.e. intelligence as process, or fluid intelligence), and (b) the outcomes of learning, namely achieving acquired knowledge and skills (i.e. intelligence as knowledge, crystallised intelligence). The authors suggested organisational psychology could benefit from models that provide a synergy between approaches. Ackerman's (1996) intelligence as process, personality, interests and knowledge (PPIK) theory is an example. Meta-theories, such as Chamurro-Premuzic's (2005) intellectual competence model, explain the interplay between personality, cognitive factors, and other influences including the real-world context.

The idea that a contextual understanding of cognitive ability might be needed (i.e. cognitive ability as applied to the workplace) raises the possibility that occupational conceptualisations of intelligence might also exist. What might be considered an indicator of ability in one job context may not be seen as so in another job context. Occupational definitions of intelligence are a possibility for future research.

Another possible direction for organisational psychology is research into work-based measurement of cognitive abilities through contextualised ability tests (Lievens & Reeve, 2012). Contextualised approaches include work-sample tests, or ability tests that have been adapted to occupational or business situations. In healthcare, for example, health literacy tests have been developed that combine contextualised cognitive ability items and applied problem-solving scenarios. In business, individuals might be asked to make inferences on the basis of business graphs (Hattrup, Schmitt, & Landis, 1992). Lievens and Reeve (2012) also emphasised the importance of how cognitive ability measurements are perceived by the stakeholders (i.e. applicants and managers within the organisation).

It is now widely accepted that general mental ability correlates with adaptive behaviours (Sternberg & Kaufman, 2012). One of the influential findings in this regard, as applied to work settings, came from Schmidt and Hunter's (1998) meta-analysis on the predictors of job performance. Schmidt and Hunter concluded that when hiring staff with no previous experience, the most valid predictor of job performance and learning was general mental ability. The next chapter of this literature review (section 2.3) discusses this finding, along with other research regarding the link between cognitive ability and performance at work.

1.2. Approaches to developing ability tests

The creation and introduction of cognitive ability tests should follow international professional standards applicable to all psychometric tests (Cohen, Swerdlik, & Sturman, 2013). Research evidence for properties such as validity (with its three aspects) and reliability are central. The same properties allow evaluation of how appropriate an ability test is for use in a particular setting (Gregory, 2011; Murphy & Davidshofer, 2005).

Reliability

High reliability is a hallmark of a sound test. Reliability is important because (1) it gives a quantified estimate of error content in a score, and (2) allows a factual statement that an individual's standing on an ability continuum falls within a sub-range with a pre-specified probability (i.e. 'confidence bands') (Cronbach, 1990; Reynolds & Livingston, 2012). Such an interpretation will indicate whether an observed test score is close to the true score or not (Ones et al., 2012). Poor test reliability can damage validation, for example, by placing a low limit on the value of a correlation coefficient found between a predictor variable and criterion variable (the 'validity coefficient') (Cohen et al., 2013). Common ways of estimating reliability of ability tests include the test-retest and parallel forms procedures, as well as internal consistency. Different methods can be used to calculate a reliability coefficient based on internal consistency (Cronbach, 1990; Murphy & Davidshofer, 2005). The split-half method correlates scores gained from one-half of the test with the other half. Cronbach's α is based on inter-item correlations and corresponds to the mean of all possible split-half coefficients. Following Nunnally's (1978) recommendation, α levels exceeding .70 are generally considered acceptable.

With abilities, Cronbach's α is by far the most commonly reported reliability coefficient; however, it is occasionally misused (Schmitt, 1996). Cortina (1993) warned that a high number of items within a test can inflate the coefficient, and that α assumes internal consistency of a unidimensional scale (i.e. it is not intended to indicate dimensionality or be applied to multiple dimensions collectively). The alternative approach to the 'true score model', the Generalisability Model of reliability (Cronbach, Gleser, Nanda, & Rajaratnam, 1972), has had limited influence on ability testing to date.

Many cognitive ability tests used in the employment sphere have excellent psychometric properties (Ones et al., 2012). A meta-analysis (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003) of reliability estimates for tests of cognitive ability found an average test-retest reliability of .83 (based on 31 coefficients). For numerical and verbal components, the average reliabilities were similar to those found for general mental ability. Other facets (e.g. spatial ability) had lower estimates, with an average reliability coefficient of .77.

Another meta-analysis (Hülsheger, Anderson, & Salgado, 2009) reported reliability coefficients ranging from .70 to .97. However, the authors did not separate out test-retest and internal consistency reliabilities. A third meta-analysis (Lang, Kersting, Hülsheger, & Lang, 2010) reported a mean reliability coefficient near .80 (over 142 coefficients).

Validity

The practice of cognitive ability testing continues to rely on the classic tripartite model of test validity. Definitions of the generic notion of validity tend to be vague, for example, 'the extent to which the test measures what it purports to measure' (Shum, O'Gorman, & Myors, 2006). Definitions of the three aspects of validity – content-, criterion-, and construct-related validity – are not replicated in this thesis (they are available in introductory works such as Cohen et al., 2013; Murphy & Davidshofer, 2005; Reynolds & Livingston, 2012). There are alternative procedures available to produce evidence for each of the three aspects of validity.

Construct validity has been addressed as the most profound layer of test validity (Cronbach, 1990). Key methods to construct-validate an ability test include convergent and discriminant validation, as well as applications of multitrait-multimethod matrix (MTMM; Campbell & Fiske, 1959). Convergent validity indicates the extent to which scores on the test co-vary with scores on other tests assessing the same construct, particularly those already fully validated. For example, scores on one measure of verbal reasoning ability would be expected to correlate with scores on another sound measure of verbal reasoning ability. Discriminant validity indicates the extent to which scores of the test under validation co-vary with scores of another test measuring an extraneous, conceptually unrelated construct.

Criterion-related validity is inherently useful for organisational psychology. Criterion-related validity is the extent to which scores on a test co-vary with an appropriate outcome (quantified by a 'criterion variable'). For example, criterion-related validation may confirm that a measure of integrity (predictor) is correlated with an outcome variable such as frequency of counter-productive behaviours at the workplace. The criterion-related validity of an ability test as a predictor of job performance is a key feature in choosing tests for a selection process.

The development of cognitive ability tests has been described as evolutionary, rather than revolutionary (Sternberg & Kaufman, 1996), implying that there have been small, gradual changes over time in how ability is assessed. Perhaps the most significant trend in the measurement of cognitive ability has been the shift towards assessing multiple facets or abilities, rather than a single intelligence dimension. This has meant that there is now greater alignment between tests and theories of cognitive ability.

The advent of Item Response Theory

The recent history of cognitive ability testing reflects the same dichotomy as the general trends in building psychometric tests. The key methodological approaches are classical test theory and item response theory.

Classical test theory focuses on applications of measurement in the realm of psychological testing, with several properties of tests being modelled. A key concern is estimating the amount of measurement error, and improving quality through improving test reliability. Central to classical test theory is the concept of reliability labelled as the 'true score model’ (Lord, 1959). The true score is the score that the individual would have obtained on a test had there been no measurement error. Error component in an observed score may relate to multitude of factors not controlled, and indeed not known, in mainstream psychometrics (Novick, 1966). The reliability of the test is assumed higher when the level of error variance in the observed scores is lower.

Classical test theory is compatible with the so-called 'classical' or traditional paradigm in item analysis (Murphy & Davidshofer, 2005). This paradigm involves estimation of several item-level parameters, mostly via statistics reliant on assumed linearity of relationships (such as the item-test relationship). The quality of individual items within a test can be evaluated using two parameters: item difficulty and item discrimination power (Cohen et al., 2013).

Item difficulty relates to the proportion of respondents who were able to answer a particular test item correctly. For items with binary outcomes (correct/incorrect), the parameter is the proportion of respondents responding correctly. Extreme values (close to zero or one), indicate that the item is not suitable for the group that the item is being tested in (Kline, 2005).

The second item-level parameter is item discrimination. This relates to how well an item discriminates or differentiates between subsets of respondents who scored high versus low on the scale overall. There are different ways of calculating item discrimination, each leading to a different statistical indicator. Corrected item-total correlation (point-biserial correlations in case of dichotomous or dichotomised items) is a common technique. Representative sampling from a tightly defined population is crucial for parameter estimation. The item difficulty and discrimination parameters can only be generalised to the population that was sampled (Kline, 2005).

Item response theory (IRT), originally inspired by Rasch models (Embretson & Reise, 2000), is considered an improvement over classical test theory (Hambleton, 1983; Hambleton, Swaminathan, & Rogers, 1991; Hutchinson, 1991; Warm, 1978). When trying to predict which response will be produced on a test item, the model takes into account a revised set of parameters. Each response on a test item is considered a mathematical function of 'person and item parameters'. For example, producing a correct versus incorrect response on an ability test item will be modelled as an outcome of a set of parameters assigned to the individual testee and parameters assigned to the item. The item parameters continue to cover difficulty of the item, but linearity of relationship between either item or person parameters, or between item parameter and the overall metric of total score, is no longer required (Embretson & Reise, 2000). The probability of a certain item-level response occurring can be estimated via a curvilinear relationship and 'item characteristic curves' (ICC) developed for each item (Shultz & Whitney, 2005).

In the language of IRT, the probability of a correct response is related to the 'latent trait' of the construct that the item is designed to assess (Hambleton et al., 1991). For ability testing, IRT exploits an item response function (IRF) to compute the probability that a person with a pre-specified level of ability will respond correctly on an item (Embretson & Reise, 2000). Individuals with higher ability have a greater chance of answering the item correctly. As well as an individual's ability, the probability of answering an item correctly will also depend on item characteristics. IRT can help re-estimate the effect of guessing a correct answer. An advantage of IRT is the possibility for a deeper analysis of item-level response patterns. Tests designed under IRT therefore have superior power to detect unusual response patterns, such as faking (Kline, 2005). IRT also allows for the more recent development of 'individually tailored/adaptive testing' where an automated system chooses subsequent test items on the basis of the particular testee's response on the preceding item. Adaptive testing abandons the classic model of standardised testing, without endangering construct validity or inter-individual comparability of measurement (Drasgow & Hulin, 1990).

CHAPTER 2: UNDERSTANDING AND PREDICTING JOB PERFORMANCE

2.1. Understanding job performance

2.1.1. The fundamental logic of using psychological information in personnel selection

A central aim of selection procedures in organisations is to identify applicants who are likely to perform well on the job. In larger organisations, a range of selection methods are often used in conjunction with each other, for example, structured interviews, psychometric testing, and work sample tests, for the purpose of identifying applicants who will be able to perform well to the requirements of the job.

Many studies have investigated the effectiveness of these methods in predicting applicants' future performance on the job. These studies typically involve analyses of criterion-related validity, estimating how well the variable (such as assessment) predicts the real-world outcome (job performance). The fundamental logic is that higher correlations between the predictor and outcome variables represent greater predictive validity.

2.1.2. Models of performance

To predict performance at work, it is necessary to both define and measure the performance domain. In simplest terms, performance can be defined as behaviour and actions that relate to the goals of the organisation (McCloy, Campbell, & Cudeck, 1994). Performance is also distinct from related concepts of work effectiveness and productivity. It is also different to outcomes, or the consequences of actions, as performance is the actions themselves (McCloy et al., 1994).

Campbell's determinants of performance and performance components

Campbell's (1990; 1993) model of performance determinants and performance components identified the common aspects of performance across jobs, and acknowledged that performance is a multi-faceted construct, rather than a broad, single component. According to Campbell, performance has three causes or determinants: declarative knowledge, procedural knowledge, and motivation. Declarative knowledge is knowledge about facts, principles and task requirements, or 'knowing what'. In contrast, procedural knowledge is 'knowing how' to perform. Motivation refers to the choice to perform, effort, and persistence. Campbell hypothesised performance to be a function of these three determinants, as shown below.

$$PC = f(DK, PKS, M)$$

(PC represents the job performance component, DK is declarative knowledge, PKS is procedural knowledge and skill, and M is motivation (McCloy et al., 1994)).

Individual differences such as ability, personality, and experience were also thought to impact on performance, but in an indirect way through influencing the determinants.

Campbell (1990; 1993) distinguished between performance determinants and performance behaviours (components). Eight performance components were identified: job-specific proficiency, non-job-specific proficiency, written and oral communication, task proficiency, demonstrating effort, maintaining personal discipline, facilitating team and peer performance, supervision/leadership, and management/administration. Of the eight components, three were proposed to be common to all jobs: job-specific task proficiency, maintaining personal discipline, and demonstrating effort. The other factors were hypothesised to vary between jobs.

2.1.3. Expansion of the performance domain

Since Campbell's (1990; 1993) model, researchers have proposed expansions to the performance domain. This has included an expansion of outcomes associated with performance and recognition of the facets within performance (Schmitt, 2014). There is greater interest in the dimensionality of performance and the specific types of performance that are important in the work context.

Typical and maximum performance

It has been proposed that understanding criterion measures of performance can be improved by distinguishing between typical and maximum performance (Sackett, 2007; Sackett, Zedeck, & Fogli, 1988). Whereas maximum performance is considered the very best an employee can do, typical performance is the employee's usual or general level of performance. Common selection procedures assess maximum performance, for example, through work sample tests. However, this may not predict long-term typical performance. It has been suggested that this error lies in the assumption that measures of maximum performance are interchangeable with measures of typical performance (Sackett et al., 1988).

Low correlations have been found between typical and maximum performance (Sackett et al., 1988), leading to the proposal that different predictors may be behind typical and maximum performance. It is thought that maximum performance is mostly related to abilities, whereas typical performance is more likely influenced by personality and motivation.

Sackett et al. (1988) also identified that supervisory ratings correlated more highly with maximum performance, rather than typical performance.

Contextual and task performance

Performance models have also been expanded by distinguishing between the types of behaviours that represent performance. One of the most influential distinctions in the literature is between task and contextual performance (Borman & Motowidlo, 1993). This is the concept that performance can include workplace behaviours besides the performance of tasks. Contextual performance describes behaviours that relate to the social and psychological work environment, rather than to the technical work environment. It contrasts with task performance, which is an individual's proficiency at performing the tasks that are formally recognised as part of the job.

Organisational citizenship behaviour (OCB) has received attention in the organisational psychology literature. Earlier definitions described organisational citizenship behaviour as discretionary, non-rewarded behaviour (Organ, 1988, cited in Bateman & Organ, 1983; Organ, Podsakoff, & MacKenzie, 2006). Later definitions (Organ, 1997) aligned organisational citizenship behaviour more closely with contextual performance (Borman & Motowidlo, 1993; Motowidlo, Borman, & Schmit, 1997). Organisational citizenship behaviour can be broken down further into altruistic behaviours (to other individuals and groups within the organisation) and generalised compliance (helping the broader organisation) (Williams & Anderson, 1991).

There is increasing interest in organisational psychology in OCB as a form of performance, with possible relationships to other variables. This may be due to the trend to look at positive aspects of a job, consistent with the field of positive psychology. The literature has indicated that measures of cognitive ability are more closely associated with task performance, whereas personality is a better predictor of OCB and contextual performance (Borman & Motowidlo, 1997; Motowidlo & Van Scotter, 1994).

A further way of looking at performance is from the opposing view, of what is considered undesirable performance. This is known as counterproductive work behaviour (CWB), and it can also be viewed as a facet of performance (Sackett, 2002). There are different types of counterproductive work behaviour, for example, CWB for self-gain or for organisational gain. Counter-productive work behaviour can therefore take a variety of forms, ranging from theft through to sabotage. It has been suggested that OCB and CWB might be different ends of a continuum; however, research suggests that these are in fact two distinct constructs, rather than a single continuum (Sackett, Berry, Wiemann, & Laczo, 2006).

Adaptive performance

Adaptive performance and how to predict it has been identified as a growing area of interest in the literature (Schmitt, 2014). Adaptive performance refers to an individual's flexibility and ability to adapt to changing circumstances and work environments. It has been proposed that emergencies tend to bring out adaptive behaviour (Pulakos, Arad, Donovan, & Plamondon, 2000), and that modern workplaces are characterised by change and uncertainty (Hesketh & Griffin, 2008).

Adaptive performance is thought to be multidimensional. Pulakos et al. (2000) proposed a taxonomy of adaptive behaviour: handling emergencies and crisis situations, handling work stress, solving problems creatively, dealing with uncertain and unpredictable work situations, learning work tasks, technologies and procedures, demonstrating interpersonal adaptability, demonstrating cultural adaptability, and demonstrating physically-oriented adaptability.

Pulakos et al. (2000) investigated the dimensionality of adaptive behaviour by investigating more than 1,000 critical incidents in 11 different organisations representing the private sector, government, and military. Participants described critical incidents, along with the behaviours representing performance in relation to these incidents. Incidents were sorted by five organisational psychologists into the proposed eight dimensions. An analysis showed that 83% of the incidents were categorised into the same dimension by 60% of the psychologists. Some jobs were thought to involve more adaptive behaviour than others, and the distribution of behaviour across the categories varied from job to job. This suggests that different types of adaptive behaviour may be required in different jobs (Pulakos et al., 2000).

Research to date into the prediction of adaptive performance suggests that the dimensions of adaptive performance are linked to predictor variables in theoretically logical ways (Pulakos, Mueller-Hanson, & Nelson, 2012). For instance, cognitive ability predicts the cognitively-oriented components of adaptive performance, such as learning new tasks and solving problems. Non-cognitive predictors are linked to interpersonal and cultural dimensions. Pulakos et al. (2012) also suggested investigating individual difference constructs, including resilience, cognitive complexity, and self-awareness in relation to predicting adaptive performance.

Adaptive performance is thought to closely align with trainability, which can be conceptualised as the ability and motivation to attain the required level of proficiency to effectively perform the job (Pulakos et al., 2012). Adaptability and trainability are arguably important to all types of jobs, given that most jobs involve training and developing capabilities to progress in the job.

2.2. Methods of assessing job performance

Subjective measures are the most common method of assessing job performance. These measures rely on another person(s) evaluating the performance of the employee. The most frequently used method is through supervisory ratings. Given the popularity of supervisory ratings, this is an area of attention in the organisational psychology literature, particularly around ways to improve the validity and reliability of this form of performance measurement (Woehr & Roch, 2012).

2.2.1. Performance measurement issues

Criterion contamination and deficiency are two concepts that apply to performance measurement. If all aspects of performance could be measured perfectly, then this would be considered the ‘ultimate criterion’ because the measured criterion would fully represent the variance in the latent performance variables (or factors). However, the measured criterion is the ‘actual criterion’, and the differences between the actual criterion and the ultimate criterion represent contamination and deficiency (Cleveland & Colella, 2010).

Deficiency refers to the actual criterion (i.e. measured job performance) missing a crucial element of performance. Contamination refers to the inclusion of information in the actual criterion that is unrelated to the behaviour that is supposed to be measured. One way to protect against contamination is acknowledging that performance is multi-faceted, and accordingly measuring specific facets of performance. Additionally, on the predictor side, it is also important to acknowledge specific facets (e.g. the components of cognitive ability).

2.3. Predicting job performance

2.3.1. Cognitive ability and job performance

Multiple studies have shown that cognitive ability predicts job performance in a variety of job settings (cf. McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Olea & Ree, 1994; Ree, Earles, & Teachout, 1994). Out of a range of different predictors, general mental ability has been consistently identified as the strongest predictor of job performance (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). It has been found to predict performance across many types of jobs of varying complexity, with stronger predictions emerging for jobs that are more complex (Hunter, 1986; Schmidt & Hunter, 1998, 2004).

Relationship to job complexity and information processing

Reeve and Hakel (2002) argued that although the relative importance of *g* varies across domains, the minimum requirement increases as the job environment becomes more complex. Research into validity generalisation is one of the strongest contributions of organisational psychology to the field of cognitive ability research.

Jobs that are more complex tend to involve heavier information processing demands, for example, dealing with unexpected situations, recalling and remembering information, solving problems quickly, learning and understanding quickly, and integrating information in order to reach a conclusion (Schmidt & Hunter, 2004). It has been suggested that complexity is the key feature in the workplace and a dimension by which jobs vary (Furnham, 2008).

However, all jobs have demands in terms of information processing, for example, problem-solving, reasoning, and decision-making. Therefore, cognitive ability has implications for meeting the challenges of all work environments. Some researchers have argued that general mental ability is more important than ever in the constantly changing and complex business world (Scherbaum et al., 2012). Additionally, all work environments involve other people, and Gottfredson (1997) noted that dealing with people tends to be complex. These types of complexities include negotiating and persuading. Gottfredson suggested that "this should not be surprising, because other individuals are among the most complex, novel, changing, active, demanding, and unpredictable objects in our environments, living and working with others is a complicated business" (p.107).

2.3.2. Empirical research into abilities as predictors

Cognitive ability as a predictor of performance

Schmidt and Hunter's (1998) influential finding came from a meta-analysis on the predictors of job performance. The conclusion was that when hiring staff with no previous experience, the most valid predictor of job performance and learning was general mental ability.

It has been suggested that 'job knowledge' mediates the relationship been cognitive ability and job performance. Using path analysis, Borman et al. (1993) investigated the relationship between ability, knowledge, experience, and proficiency, and the effect on ratings in a military sample. Ability was found to have a greater effect on job knowledge than on task proficiency. Schmidt and Hunter (1992) investigated causal models between general mental ability, job knowledge, job performance (via a work sample), and supervisory ratings of performance in both military and civilian groups. The authors found that the effect of general mental ability on job performance was indirect and occurred through the acquisition of job knowledge. Similarly, the main determinant of supervisory ratings was job knowledge, rather than work sample performance.

Individuals high in general mental ability also tend to acquire substantial declarative and procedural knowledge (Ones et al., 2012). Kuncel, Hezlett, and Ones (2004) argued that higher cognitive ability relates to faster knowledge acquisition across settings, including training and on the job. Ones et al. (2012) looked at the results of almost 80 meta-analyses that investigated the relationship between cognitive ability variables and training success. The meta-analyses covered a wide range of occupational groups and jobs ranging in levels of complexity. Two findings were evident: firstly, that cognitive ability related to training success, and secondly, the relationship was higher when the knowledge to be acquired through the training was more complex. Given the association of general intelligence with learning, it could be argued that higher cognitive ability also reflects higher trainability at work.

Jobs that are considered high prestige tend to require more training. This observation led some to suggest that perhaps training was responsible for these jobs having higher status. Gottfredson (1997) disagreed, and argued that these high prestige jobs also required continual learning, updating of knowledge, and in general the exercise of high ability. Therefore, individuals in these jobs required learning even after the job-specific training was completed. Gottfredson argued against the possibility that training can counteract for differing levels of cognitive ability between employees. The predictive value of training drops when workers have more experience, but cognitive ability remains a significant predictor.

Facets of ability as predictors of performance

Lievens and Reeve (2012) suggested that investigating the correlates of specific abilities is one important area for organisational psychology to focus on in the future. There is evidence that specific factors can be of importance in addition to *g* (cf. Park, Lubinski, & Benbow, 2008). This does not mean the discrediting of *g*, but investigating how the two can work together. However, the contribution of *g* and specific abilities to the prediction of job performance is debated. Furnham (2008) concluded that measuring very specific abilities does not add much in terms of incremental advantage over general mental ability when predicting job performance.

Group differences

Psychometric tests of cognitive ability have at times been controversial. Some studies have revealed group differences in mean scores raising the possibility that the use of psychometric instruments to assess cognitive ability during selection might disadvantage particular groups and lead to adverse impact. Roth et al. (2001) undertook a meta-analysis into ethnic group differences on cognitive ability tests in the United States and found significant differences (with minority groups scoring lower), particularly for general cognitive ability more than specific facets. Guenole, Englert, and Taylor (2003) looked at differences in cognitive ability scores in a New Zealand sample and found significant differences between ethnicities, including that Māori scored lower than New Zealand European on measures of verbal reasoning and numerical business analysis, but there were no differences in general numerical reasoning. The authors suggested strategies for minimising adverse impact. These included within-group norming, different cut-off scores by ethnicity, placing more emphasis on non-cognitive performance constructs (such as interpersonal ability), and using methods that have shown no ethnic group difference, such as structured employment interviews (Guenole et al., 2003).

Gender differences in general and specific cognitive abilities have been the topic of a large amount of research, but the findings have been inconsistent. A report for the American Psychological Association by a task force of researchers (Neisser et al., 1996) claimed that psychometric tests are constructed in such a way that there will be no differences between male and female respondents. This view is consistent with findings that have found negligible gender differences in general cognitive ability (Colom, Juan-Espinosa, Abad, & García, 2000). Other studies have reported there may be gender differences in specific abilities (Neisser et al., 1996). The common gender differences reported were that males tend to outperform females on assessments involving mathematical reasoning (cf. Benbow, 1988) and mental rotation (cf. Linn & Petersen, 1985), whereas females tend to do better in tasks involving verbal reasoning (cf. Wai, Cacchio, Putallaz, & Makel, 2010). Johnson and Bouchard (2007) argued that gender differences in reasoning abilities were only observed when the effects of *g* were controlled for, suggesting that general ability tended to mask gender differences in specific abilities.

Other meta-analyses have found no gender differences in specific abilities. Hyde and Linn (1988) conducted a meta-analysis looking at gender differences in verbal reasoning ability.

Overall, females showed slightly better performance on verbal reasoning; however, the effect size was small (*d* = .11) to the extent that the authors concluded that gender differences in verbal reasoning ability no longer existed. Hyde, Fennema, and Lamon (1990) conducted a meta-analysis of more than one hundred studies that looked at gender differences in mathematical ability in different age groups. Overall, males performed slightly better but the effect size was small (*d* = .15). In the general population, females performed slightly better, although the difference was negligible (*d* = -.05). The authors concluded that any gender differences in mathematical ability were small.

In relation to the higher representation of men in scientific and mathematical occupations, Spelke (2005) reviewed the literature and investigated possible theories for the disproportionate numbers, including the possibility that males may have greater natural abilities in spatial and mathematical reasoning. Based on the research from different ages (from birth through to adulthood), Spelke concluded that males did not have a natural aptitude for maths and science, and that cognitive ability was not the reason for lower representation of females in scientific occupations (suggesting that other reasons, e.g. social and cultural were more likely). Another argument was that males have greater variability in cognitive ability scores, and as a result, tended to have higher representation in the extreme higher scoring end, for example, in the top 10% (Hedges & Nowell, 1995).

Other considerations

Although the evidence points towards cognitive ability as a strong predictor of job performance, caution is needed about drawing conclusions about individuals based on single tests (Sternberg, 2014). Although cognitive ability is a valid predictor of job performance, it is not the only predictor (Reeve & Hakel, 2002). Therefore, cognitive ability tests are 'incomplete predictors' of success and must be supplemented with other tests. Other worker characteristics also affect job performance and can at least co-determine outcomes. Personality traits are one such example, and it has been acknowledged that conscientiousness adds usefully to predictions of job performance (Furnham, 2008). Other individual differences (e.g. motivation) and contextual factors can also influence the extent to which an individual's ability translates into performance outcomes. Other individual differences are also important at work, for example, an individual's personality and preferences, and how these fit within an organisational culture or climate (Murphy, 1996).

There are also issues that must be taken into account when measuring cognitive ability, including reliability and validity. When correcting for range restriction and attenuation, the correlation level is around .3, meaning that cognitive ability tests predict around 10-30% of real world performance (Sternberg, 2014).

2.3.3. Validity in predicting job performance

There are a number of issues associated with validity in predicting job performance. One of these relates to the generalisability of any findings. In predicting job performance, generalisability refers to the extent that the criterion-related validity in the sample corresponds to the target population. Several factors are known to influence generalisability, including the sampling method and characteristics of the sample population. Researchers typically calculate correlations in a sample population. However, the sample (observed) correlation can vary from the population correlation in a number of ways, including restriction of range.

Range restriction

Range restriction is commonly encountered during validation of selection instruments using a sample of existing employees of the organisation, for instance, when determining the relationship between an assessment tool and job performance. This can occur when employees were earlier selected into the organisation based on scores on one of the predictor variables (for example, due to a cut-off score in a selection test). This means that all applicants scoring outside the cut-off score (usually below) are excluded from the sample, as they were not hired in the first instance.

The effect of range restriction on correlations is well documented and takes different forms depending on the selection processes by which selection takes place and the reasons for the truncated sample (Sackett & Yang, 2000). Accordingly, there are different methods for correcting for range restriction. It is not always appropriate to correct for range restriction, and whether to or not depends on the population the researcher wishes to generalise the findings to (Sackett & Yang, 2000). For instance, if the sample population and the population of interest were the same, then correcting for range restriction would not be appropriate. An example would be a validation study predicting the usefulness of cognitive ability in predicting performance of existing staff. However, if the research question was to predict the future performance of job applicants, then the sample population and population of interest would be different, and correction for range restriction would be appropriate. The implication for validation studies is that researchers need to clearly define the population of interest, be aware of the causes of selection bias, and carefully consider the appropriate facets that may interact to cause range restriction (Sackett & Yang, 2000).

Characteristics of the sample

A further threat to the external validity (generalisability) of any findings relates to the characteristics of the sample population, particularly how these characteristics correspond to the population of interest. If the criterion-related validation study was undertaken in a sample of all male participants, the findings could be expected to generalise to the male population rather than female population. As individuals can vary in a multitude of ways, researchers usually use sampling strategies to randomly select participants for the sample. This is one way to increase the generalisability of the findings.

With respect to the validation of psychometric tests, the composition of the sample group is an important consideration. Researchers should validate a test with participants as similar as possible to the respondents with whom the test will be used in practice. For example, if a test will be used with New Zealand job applicants, then the test should be validated with a sample from New Zealand. This is also important for developing appropriate norms for use with the test during selection procedures.

The importance of the sample also applies to occupational groups. For instance, a test that will be used in the prediction of performance for engineers should be validated with a sample of engineers, rather than an occupationally dissimilar group. This also allows for the development of local norms at the level of the occupational group.

By considering factors such as these, researchers increase the validity of the test within the specific context it will be used. Taking steps to ensure validity also helps to ensure that respondents or job applicants are participating in fair selection procedures, and the organisations are able to justify, with evidence, their use of a particular selection tool within a particular context.

CHAPTER 3: CORRECTIONAL WORK AND PREDICTING PERFORMANCE

3.1. Introduction to the corrections system in New Zealand

The Department of Corrections is responsible for administering New Zealand's corrections system. Forming part of the government's wider justice sector, Corrections works alongside the Ministry of Justice, Police, New Zealand Parole Board, and the Courts, to contribute to a 'safe and just society' (Department of Corrections, 2014). Corrections' priorities are to improve public safety and to reduce reoffending. To support these aims, the Department manages sentence compliance and ensures that prisoners, parolees, and other community-based offenders comply with the conditions of their sentences and orders. The Department provides rehabilitation and reintegration interventions to target offenders' risks and needs, with the aim of increasing the likelihood they will go on to lead crime-free lives.

As one of New Zealand's largest government departments, Corrections employs more than eight and a half thousand staff, most of whom work directly with offenders at prison and community corrections sites. These 'frontline' or 'offender-facing' employees include approximately three thousand corrections officers, who work with sentenced and remanded prisoners at eighteen prison sites throughout the country. Other offender-facing employees include probation officers, who work with offenders on community-based sentences and orders, programme facilitators, case managers, psychologists, offender employment instructors, and community work supervisors.

Offenders have contact with staff in these various roles as they move through the corrections system. However, offenders transitioning through the prison system have more interactions with corrections officers than any other role due to the frequent nature of the day-to-day contact. There is increasing interest in the role of frontline officers, not only in managing safety and security functions, but also in their potential to contribute more directly to offenders' rehabilitation.

3.1.1. Correctional philosophies and principles

Correctional systems, policies and practices are influenced by theories and philosophies about criminal justice. These philosophies have varied between jurisdictions, and have evolved with political, legislative, and social changes. Historically, corrections systems were viewed as systems to administer punishment. The contemporary view is that corrections systems must also contribute to changing behaviour through addressing the causes of an individual's offending (i.e. rehabilitation). There is a tendency in the literature to dichotomise philosophies about criminal justice as punitive versus rehabilitative; however, in most jurisdictions corrections systems involve elements of both punishment and rehabilitation.

The balance of punishment and rehabilitation in corrections systems has long been a topic of debate. Throughout the twentieth century, significant changes took place in public and professional opinion around correctional approaches. Correctional systems evolved to recognise rehabilitation as an accepted correctional goal, and new correctional approaches beyond punishment were explored (Wicks, 1974).

However, by the 1970s rehabilitative approaches were being questioned, particularly in the United States, which was experiencing rising crime rates and prison overcrowding. Reviews of treatment evaluation published during the mid-1970s were unable to provide evidence that rehabilitation significantly contributed to reducing reoffending (Lipton, Martinson, & Wilks, 1975; Martinson, 1974)[1]. These publications were influential, and in general the findings were taken to mean that 'nothing works' in relation to changing offending behaviour. Combined with the political climate and prevalent ideologies at the time, an emphasis was given to stronger penalties and less investment into rehabilitation programmes (Cullen & Gendreau, 2001; Gendreau, 1996).

Some researchers challenged the pessimism of Martinson's (1974) review and disputed the opinion that rehabilitation was futile. Subsequent reviews, including Palmer (1975) and Gendreau and Ross (1979), queried the quality of Martinson's review and presented a more balanced argument, providing explanations for the failure of some interventions, and evidence to support the success of other interventions in preventing reoffending. In the 1990s the volume of literature on correctional rehabilitation grew. This research became known as the 'what works' or 'evidence-based practices' literature, and it represented a shift away from the view that 'nothing works' in rehabilitating offenders (Department of Corrections, 2009).

A number of literature reviews and meta-analyses have since investigated the characteristics of effective rehabilitation (cf. Andrews, Zinger, et al., 1990; Lösel, 1995; Mulvey, Arthur, & Reppucci, 1993). It became standard for studies to measure the effectiveness of correctional treatment in relation to achieving reductions in recidivism (reoffending) rates. The literature consistently identified that *some* rehabilitation interventions were effective for *some*

[1] Martinson (1974) analysed 231 programme evaluation studies published between 1945 and 1967 and found no effect of rehabilitation on recidivism. The 1975 publication (Lipton, Martinson & Wilks) is a larger report of the findings.

offenders, and the interest shifted to identifying the specific factors that contributed to the effectiveness of rehabilitation programmes (Antonowicz & Ross, 1994).

Influential principles concerning the management and rehabilitation of offenders include the principles of risk, needs, and responsivity. These principles, which come from the body of literature known as the 'psychology of criminal conduct', provide a framework for improving rehabilitative outcomes (Andrews & Bonta, 2010; Andrews, Bonta, & Hoge, 1990). The risk principle states that higher risk offenders should receive the most intensive and greater amount of treatment. The needs principle states that treatment should target offender's criminogenic needs – or their offending-related factors that can influence recidivism. The third principle, responsivity, states that treatment is more effective when it is undertaken in a cognitive-behavioural treatment programme, and when the style of service delivery is matched to the offender's learning style (Andrews & Bonta, 2010).

Correctional programmes in New Zealand followed a similar history as the United States and British models, and as such, have been influenced by the 'what works' debates and literature (Department of Corrections, 2007). Current rehabilitation approaches in New Zealand's correctional system are informed by an evidence-base about the effectiveness of treatments and interventions in changing offending behaviour (Department of Corrections, 2009). Evidence-based principles underpin correctional practices, ranging from risk assessment, managing offenders proportionate to their risk profile, the design of programmes and interventions, and matching rehabilitation programmes to offenders' assessed needs.

There is a large amount of literature on the principles of correctional rehabilitation programmes. However, there is also evidence that applying these principles in everyday correctional work can achieve positive outcomes. A Canadian study by Bonta et al. (2011) trained a group of probation officers to incorporate risk, needs, and responsivity principles into their everyday interactions with offenders. Controlling for age and criminal history risk factors, the rate of recidivism two years later was 15% lower for offenders on the caseloads of the trained probation officers, compared to offenders assigned to probation officers who had not been trained in the principles. Studies such as this highlight the contribution that frontline staff can make in achieving positive correctional outcomes. Many areas of correctional work are moving in the direction of officers having a more direct role in facilitating behaviour change (Bourgon, Gutierrez, & Ashton, 2011).

3.1.2. Legislative and institutional environment

Corrections operates within a legislative environment, and legislation governs many aspects of the day-to-day work carried out by frontline employees. The Corrections Act 2004 covers the responsibility of the Department for administering sentences and orders in a safe, humane and effective way; operating correctional facilities; rehabilitating offenders; and providing information to the Courts and Parole Board.

The Corrections Act 2004 also establishes the statutory roles and responsibilities of staff within the corrections system, including corrections officers, and along with the Corrections Regulations 2005, defines what powers they have and under what conditions they can be used. These powers include the use of force and restraints, searching prisoners, visitors and other staff, and drug and alcohol testing. Additionally, the Corrections Act 2004 protects the rights of prisoners' by outlining procedures for making complaints.

Human Rights and the ethical treatment of offenders is an important part of Corrections' institutional environment. The Corrections Act 2004 specifies that prisoners must be managed in a way that is humane, as well as safe, secure, and effective. The New Zealand correctional system also abides by international regulations for prisons, the foremost being the United Nations Standard Minimum Rules for the Treatment of Prisoners, which in turn informs the Corrections Act 2004.

A defining characteristic of the prison institutional environment is its organisational structure. Prisons are traditionally hierarchical and have a 'chain of command'. Although the term 'paramilitary' is often used in the literature, most of the literature originates in the United States. Officers wear a uniform with epaulettes representing their rank. Although the chain of command is present in the prison daily routine, it is most crucial in emergency response scenarios, where safety and security are paramount.

Prisons use a 'unit-based' management model. The prison is made up of a number of units, and each unit houses prisoners usually of a specific category or security classification (e.g. high-risk prisoners, 'at risk' prisoners, remand prisoners). Officers are generally assigned to particular units. Dividing the prison into units has a number of benefits, including creating smaller, more manageable work environments, creating closer networks and communication between staff members, and allowing staff members to build better knowledge of the prisoners in their unit (Hobbs & Dear, 2000).

From an operational perspective, prisons are traditionally prescriptive, regimented and rule-driven environments, relying heavily on policies and procedures. This prescriptive environment is a feature of correctional systems internationally. The correctional culture in the United States has been described as hierarchical, with detailed rules and regulations, top-down communication and strong military overtones. There is also uniformity of dress and conformity to rules and regulations (Farkas & Manning, 1997). Although some of these characteristics also apply to the New Zealand corrections system, it is not considered militarised as in the United States.

This classical model of corrections organisations has been critiqued. Gilbert (1997) argued that prisons are not military organisations, but instead they are primarily social service organisations. Unlike military organisations, the most important decisions are made by the lowest-ranking officers, through their daily routines and interactions with the prisoners. It was suggested that communication and interpersonal skill are more important than rank or structure in the prison's effectiveness and maintaining safety and security (Gilbert, 1997).

Corrections and police have elements in common; for example, they are both rule-centred and hierarchical systems (Farkas & Manning, 1997). Both roles deal with unpredictability and risk; however, it is somewhat less for corrections officers as they are confined to one physical environment and deal with mostly the same individuals. Due to the similarity, some of the correctional literature makes comparisons with the police organisational environment.

3.2. The corrections officer role

3.2.1. The nature of the job

The literature uses a variety of terms to refer to corrections officers, and Lombardo (1989) described how the language has evolved along with perceptions of what the job entails. In the United States, the term *correctional officer* replaced earlier names of guard and warden during the movement to 'professionalise' the work of corrections staff.

The different facets to the corrections officer role have been identified and debated in the literature. Lombardo (1989) termed the different aspects of the corrections officer role as the 'people worker' and the 'bureaucrat'. As a 'people worker' the officer works with prisoners on a human level and focuses on the prisoners as individuals. This includes finding ways to influence prisoners' compliance. As the 'bureaucrat' the corrections officer works within the complex organisational system, follows formalities and rules, and views the prisoner less as an individual and more generally as an offender.

Job descriptions identify that the corrections officer role has two main functions; the first covers custodial duties focused on safety and security, and the second includes the rehabilitative and care components.

Safety and security functions are vital in a prison in order to protect the prisoners, staff, visitors, and the public. Corrections officers must account for the whereabouts of prisoners at all times. Routine security tasks include the prison muster (count of prisoners), patrols, and cell checks. Checking for contraband (banned items such as cigarette, drugs, cell phones and potential weapons) is also part of the routine, and officers are responsible for searching prisoners and their cells.

In the unit management model, corrections officers are considered the first source of support for prisoners. This includes assisting or referring on prisoner's welfare matters, and assisting with practical matters. In a study of staff and prisoner interactions in a Western Australian maximum-security prisoner, Hobbs and Dear (2000) found that prisoners were more willing to seek practical assistance than emotional support from officers. This is consistent with the traditional view of a corrections officer having a custodial rather than rehabilitative role. Although there are several potential explanations for this finding, for example, that the prisoner views the officer as an authoritarian or disciplinarian figure, it does highlight the complex staff-prisoner interface, and how this is influenced by attitudes and beliefs of both parties.

Working with prisoners – challenges and working with risk

It is useful to examine the unique characteristics of the correctional work environment that influence the nature of the job. The challenges are well recognised in the literature. Corrections officers work with a mandated confined population, and the term 'involuntary client' is sometimes used to describe the individuals with whom the correctional system deals with. Imprisonment involves the restriction of everyday freedoms, and prisoners are required to comply with the prison routine, policies and procedures, attend programmes, and be accounted for at all times. Correctional work has been described as adversarial and conflicting in nature, with officers and prisoners in conflicting roles (Farkas & Manning, 1997), and it is generally assumed that prisoners are reluctant, unwilling and unmotivated to comply. However, although these descriptions are a common theme in the literature, there is variation in how prisoners adapt and respond to the prison environment and staff.

Farkas and Manning (1997) described correctional work as characterised by uncertainty and unpredictability. Most of the time the prison routine is monotonous and consists of core tasks and routines, but the possibility of an incident always exists; these can range from minor disturbances and rule breaches, through to less common incidents such as assaults, hostage taking and riots.

Aside from potential physical harm, there is the risk of being psychologically manipulated by prisoners. This point was made by Wicks (1974) who advised correctional officers to maintain a degree of attentiveness due to the fact they deal with antisocial and psychopathic individuals. Wicks reminded corrections officers that while they needed to treat the prisoner as a human being, they must also be alert to manipulation attempts. Officers may be coerced into such activities as bringing in contraband for a prisoner through to becoming involved in gang-related or criminal activities, for example.

Corrections officers work in an environment that requires constant awareness and response to changes in risk. Risk in the corrections context can have a range of meanings; however, offender risk generally refers to the likelihood of an offender committing further offences. A prisoner’s risk profile is influenced by both static and dynamic factors. Static risk scores are calculated based on the characteristics of the offender and their criminal history that cannot be changed, for example, present age, age at first conviction, number of convictions, and seriousness of offence.

A prisoner’s risk profile is also influenced by dynamic risk factors. These prisoner characteristics can potentially change. When directly linked to patterns of offending these are known as ‘rehabilitative needs’ or ‘criminogenic needs’. Examples are substance use, antisocial associates, and propensity for violence. There is increasing focus on dynamic risk assessment in prison, and this reflects evolving thinking around the characteristics of prison-based offenders. Unlike the community corrections (probation) environment, prisons were traditionally viewed as more static environments with less potential for constant risk changes. However, there is now increased recognition that risk can and does change during a prisoner’s sentence and needs to be assessed and managed continuously (cf. Atkinson & Mann, 2012; Gendreau & Keyes, 2001).

Corrections officers are able to provide information about the prisoners’ recent behaviour and circumstances to help inform other correctional roles in their assessment and planning for individual prisoners (cf. Atkinson & Mann, 2012). Corrections officers are also instrumental in managing changes in risk on a practical day-to-day level, and they help to support the prisoner's rehabilitation plans (for example, by helping to ensure prisoners attend scheduled programmes).

Stress and role conflict

The international research literature on corrections officers largely focuses on the impact of the correctional environment on employees. There is much research on the challenges and demands of the role that arise from working in the prison system, for example, the predictors of job stress and staff turnover for officers. This is an important area given the association of stress with unsafe practices, high turnover, decreased effectiveness, as well as negative personal and social outcomes, such as work-family conflict.

Finney et al. (2013) reviewed literature on the relationship between job stress, burnout, and organisational stressors in correctional officers. By analysing research articles from a number of different countries, the researchers identified a number of organisational stressors in the correctional environment. These included stressors that were intrinsic to the job, which included working overtime, prisoner overcrowding, under-staffing, training, and limited resources. Workload was significantly associated with stress, as well as lack of resources.

Role problems, particularly role conflict, were associated with correctional officer stress (Finney et al., 2013). The authors suggested that this was due to the requirement of corrections officers to act professionally within the strict hierarchies and regulations of the environment, whereas managing prisoners requires flexibility. For instance, officers often use informal interactions with prisoners that are not prescribed by the strict rules and regulations. This suggestion is consistent with the observation that corrections officers exercise a reasonable amount of discretion, for example, when solving prisoners' problems and dealing with rule breaches. Gilbert (1997) argued that the discretionary power of corrections officers is under-recognised, and this discretionary power is influential in shaping the prisoner-officer dynamics.

Role conflict is also a recurring theme in the correctional literature. In the New Zealand prison environment, it was recognised that tension could arise from officers needing to manage the three main accountabilities of their roles: security, care, and rehabilitation (Young, 2013). This refers to the requirement of officers to manage prisoners in a way that manages their risk (security), while also responding to the prisoner's needs (care), and contributing to their rehabilitation. Young (2013) suggested that tension arises from the need to apply all three accountabilities, and knowing when to focus on one accountability more than another.

Finney et al. (2013) identified that the organisational structure and climate had the most consistent relationship with correctional officer stress and burnout. These stressors included unclear goals and roles, lack of decision-making ability, and lack of organisational support. A limitation of the review by Finney et al. is that six out of the eight studies reviewed were from the United States, and the results would not necessarily generalise to New Zealand prisons.

Job satisfaction

The more positive aspects of the role and what makes the job rewarding have to date received less attention, and less is known about the predictors of job satisfaction for corrections officers. Job satisfaction is important as there is a relationship between job satisfaction and turnover (Lambert, Hogan, & Barton, 2001), and turnover is known to be high for corrections officers (Udechukwu, 2009). However, findings regarding the relationship of job satisfaction to job performance are mixed, and relationships are often lower magnitude (Judge, Thoresen, Bono, & Patton, 2001).

Yang, Brown, and Moon (2011) investigated job satisfaction in a sample of 400 South Korean corrections officers. Using survey methodology, the researchers studied job satisfaction in relation to the five determinants on the Job Descriptive Index (JDI): pay, promotions and promotional opportunities, co-workers, supervision, and the work itself. The results indicated that only 6.2% of corrections officers were satisfied with their jobs, 52.9% were somewhat satisfied, and the remaining 40% were either somewhat dissatisfied or not satisfied at all. The five JDI dimensions were found to have varying effects on job satisfaction. Pay, promotion, and co-workers were significant at the $p < .01$ level, the job itself significant at the $p < .05$ level, and supervision at the $p < .10$ level. Despite overall low levels of satisfaction, officers who believed they were better paid, had more promotion opportunities, and had positive co-worker and supervisor relationships were more satisfied. With respect to the job itself, officers who believed the job was more challenging, had higher levels of job satisfaction.

Beliefs and attitudes about correctional roles

Identifying the beliefs and attitudes of correctional officers and the relationship to various aspects of the job is another area of research in the correctional literature. Farkas (1999) suggested that the underlying beliefs and attitudes held by corrections officers set the tone for staff-prisoner interactions. Farkas sought to investigate the attitudes of correctional staff and how they are influenced by individual characteristic variables, as this has implications for the selection of officers. The study was carried out using questionnaire methodology in two

County Corrections Houses in the United States. The questionnaire items assessed officer attitudes in relation to four areas: preference for counselling roles, punitive orientation, social distance with offenders, and concern about corruption of authority.

The results indicated that most officers did not have a punitive orientation, and most showed strong support for rehabilitation (e.g. disagreeing with the statement that "rehabilitation programmes are a waste of time and money"). However, most did not think this was part of their role. A significant proportion of the sample was concerned about corruption of authority (e.g. agreeing with the statement "a personal relationship with inmates invites corruption"), which is consistent with other literature that identifies manipulation as a challenge of the working environment.

The study also identified a number of individual and work variables predicting corrections officer attitudes. Officer age was associated with support for rehabilitation, suggesting that age and experience may have produced officers with a more flexible outlook on their role. Alternatively, officers without this support for rehabilitation may drop out of the role sooner (Farkas, 1999). A similar pattern was found for seniority, with officers working longer on the job having more supportive attitudes towards rehabilitation. Female officers were found to be more inclined towards counselling roles, yet also held more punitive attitudes, and were more concerned about corruption of authority.

Farkas (1999) suggested that female officers may have had trouble in finding the appropriate approach for the role, and that this was in part due to working in male correctional institutions and in a traditionally male role. Consistent with previous research, role conflict was found to be a significant predictor variable, and officers with high role conflict did not support rehabilitation. At the same time, they did not hold punitive attitudes, but they did have a greater desire for social distance. Farkas suggested that these officers were likely experiencing confusion about what was required of them, particularly given the political and ideological climate of 'get tough' that was present in the United States at the time of this study.

Counter to previous research, was the inverse relationship of job satisfaction to preference for counselling roles. Farkas (1999) explored this finding in more depth using open-ended questioning. It was found that extrinsic aspects of the job (such as pay, job security, benefits) were related to job satisfaction, whereas intrinsic factors (such as the challenge of the work, or the chance to help offenders) were not related to job satisfaction. Farkas proposed that this may have been due to role conflict, and suggested that the nature of the work itself was not satisfying.

There is a large amount literature on role conflict in corrections officers, and the characteristics of officers and the working environment that can lead to role conflict. Studies have also investigated the relationship between role conflict and outcome variables including stress, turnover, punitiveness, burnout, job dissatisfaction, and negative attitudes towards rehabilitation. Philliber (1987) suggested that role conflict could be due to the transition from guarding duties to a more encompassing role that includes rehabilitative components.

Corrections officers and expansion of responsibilities

The idea that corrections officers contribute to rehabilitation is not a new concept. Wicks (1974) suggested that prison security staff (i.e. corrections officers) should be trained in the basic principles of psychology. Wicks made this suggestion based on the observation that a lack of understanding of the rehabilitation process by security staff could lead to rehabilitation staff and security staff working against each other. Wicks also noted that corrections officers have greater contact with prisoners than psychologists and it is “desirable to make the security personnel part of the therapeutic community” (p. 153).

Farkas (1999) discussed efforts in the United States to shift the role from a custodial to a human services professional. Corrections officers have direct and prolonged contact with offenders, and are therefore an important source of influence (Farkas, 1999). There has been interest in the professionalisation of the corrections officer internationally, recognising the potential for frontline staff to have a role other than simply custodial duties.

Antonio, Young and Winegard (2009) identified ‘diffusion of responsibility’ as an issue in correctional settings. Diffusion of responsibility can occur when multiple people are involved in an activity, and the sense of responsibility of each person diminishes. For example, as employees in different roles have contact with prisoners, corrections officers may not see rehabilitation as their role. However, officers spend the most time with offenders, and their response to situations could either reinforce or impede treatment efforts undertaken by rehabilitation staff (Antonio et al., 2009).

Antonio et al. (2009) explored how a training programme for corrections officers could influence this phenomenon in a prison setting. In a US prison, staff were trained in reinforcing positive behaviour. The training programme clarified responsibilities in role-modelling good behaviour and correcting unacceptable behaviour in the prison. An attitude survey administered post-training identified that corrections officers perceived themselves as being the most responsible for promoting a positive social environment in the prison (Antonio et al., 2009).

Increasing the potential for intrinsic rewards in a job is known as job enrichment. There is an association between job enrichment and job satisfaction, and Hepburn and Knepper (1993) explored whether defining and expanding the corrections officer duties and responsibilities could result in increased job satisfaction. This redefinition was from a 'correctional security worker' to a 'human services worker' (Hepburn & Knepper, 1993). The study was carried out at the Arizona Department of Corrections, which was unique at the time as there was a two-track career path for officers. All officers completed the same seven-week initial training, but correctional officers could then choose to follow a correctional counselling or treatment track (and become correctional programme officers), or correctional custody track.

The results showed that the correctional programme officers had a significantly greater sense of authority over prisoners, and a significantly lower sense of role strain than correctional service officers. The correctional programme officers also found greater intrinsic and extrinsic rewards in the job and had overall higher job satisfaction. Hepburn and Knepper (1993) suggested that the wider responsibilities of the correctional programme officers increased the intrinsic rewards of the job and reduced the role strain. However, due to the self-selection of the corrections officers into the different tracks, it was possible these individuals differed in some way in attitudes and characteristics, independently of the working environment.

Whitehead (2014) discussed initial findings of the success of therapeutic communities in New Zealand prisons. The 'therapeutic communities' discussed are particular units that are set up for sub-types of offenders, with the view that the prison (or unit) itself is the programme, for example, Special Treatment Units, Child Sex Offender Units, and Drug Treatment Units. Statistically significant reductions in recidivism were found for all therapeutic community programmes. Whitehead also discussed the challenges of setting up these communities within prison, including the competing frameworks of custodial and treatment staff. For example, custodial staff found it difficult to reconcile safety and security functions with the requirements of the treatment unit, and treatment staff found the custodial routine as authoritarian and punishment-oriented. At the time this study was published, the Department was strengthening its offender-centric focus by adopting a multidisciplinary, collaborative 'One Team' approach to offender management. This is leading to greater understanding on the value of each team member's role in offender rehabilitation, as well as an increased focus on the motivating/influencing components of the corrections officer role.

3.2.2. Job requirements and expectations

Corrections officers are responsible for the day-to-day supervision of prisoners and maintaining the security of the prison. Custodial duties include managing security, locking and unlocking prisoners, searching prisoners and cells for contraband, escorting prisoners from one part of the prison to another, and patrolling the perimeter to maintain security. Prisons operate all hours and shift work is characteristic of the prison environment. The roster requires officers to work evening, night, and weekend shifts, as required.

Corrections officers are expected to be the first source of support for prisoners in their unit to assist with queries, settle disputes, resolve minor complaints, provide information, and refer on queries to other staff. In discussing what is required of corrections officers from a 'human services' perspective, Hepburn and Knepper (1993) identified high levels of informality, friendliness, and interpersonal skills. Although officers are encouraged to build an effective working relationship, at the same time, they must be careful to maintain professional boundaries.

Integrity is required of corrections officers. As identified earlier in this review, manipulation by prisoners is a challenge of working with the offender population. Officers must maintain integrity in their activities inside and outside of work. Integrity is also required in relation to the professional and ethical aspects of the job. For example, officers have access to sensitive information about prisoners, and must exercise judgement in privacy and confidentiality.

The prison population includes a wide range of characteristics and needs. More than one quarter of the prison population in New Zealand is affiliated to a gang (Department of Corrections, 2013). The prison population includes young (youth) offenders, and many prisoners have mental or physical health issues and backgrounds of substance and alcohol abuse. Suicide and self-harm is a risk in a prison, and officers need to be aware of signs and changes in behaviour and respond accordingly if a prisoner is at risk. Getting to know the prisoner's characteristics is emphasised from both a safety and security perspective (i.e. knowing the prisoner's risks), and proactively working to minimise these, and from a rehabilitation perspective (i.e. knowing which needs are targeted in the offender's rehabilitation plan).

Prisoners come from a range of cultural backgrounds. Officers are required to maintain awareness of how culture and religion may impact on the management of the prisoner and how to manage requirements within prison policy. Effectiveness with Māori is a particular area of focus given Māori over-representation in the New Zealand prisoner population (Department of Corrections, 2013).

Strong communication skills are required, and this is important for safety and security, as well as in the day-to-day interactions with other staff and prisons. Officers use tactical communication to prevent and de-escalate situations. Tactical communication involves understanding the emotional bases of anger, body language and tone, and using verbal strategies to reason with prisoners and defuse anger to prevent situations escalating to violence.

Corrections officers are required to have a good understanding of human behaviour and alertness to behavioural cues which indicate 'something is going down', such as a potential assault, escape, conflict, suicide or other disturbance or incident (Farkas & Manning, 1997). If they are the first on the scene in an event, an officer is expected to follow prescribed steps in radioing for support, managing the incident, and preventing harm to others.

Entry requirements and training

The initial training for corrections officers in New Zealand consists of classroom training and on-job training. Officers come from varying backgrounds, and formal qualifications are not a pre-requisite. The initial training is designed to cover the core skills and knowledge to produce an operational officer. Corrections officers then complete on-going training to learn new skills and maintain competence in core skills.

3.2.3. Job activities and behaviours in New Zealand prisons

The job description for corrections officers working in New Zealand identifies key areas of accountability: security, supporting offender management, compliance, health and safety, administration, personal development, and teamwork (Department of Corrections, 2012). Each of these will be examined in turn.

The security accountability includes duties relating to the daily routine of the unit, managing sentence compliance (i.e. containment of prisoners), ensuring that offenders' behaviour is in line with the required code of behaviour and legislative requirements, and activating emergency procedures when required. The duties that come under the category of 'security' are closely related to those relating to the 'compliance' accountability. Corrections officers are required to personally comply with prison policies and procedures, as well as legislative requirements, when managing offenders.

The health and safety accountability covers a broad range of duties, including identifying and reporting health and safety hazards, complying with health and safety policies and legislation, taking precautions (e.g. wearing gloves when searching), through to providing first aid and basic life support.

The offender management accountability includes the human-oriented duties and rehabilitative component of the corrections officer role. Corrections officers are required to encourage offenders to complete rehabilitative programmes. Corrections officers are expected to apply 'active management' techniques in their work with offenders. This includes providing motivational feedback, role-modelling pro-social behaviours, and providing feedback on what is good or unacceptable behaviour. Communication plays a large part in offender management. Corrections officers are expected to communicate with other staff to support managing the offender through their sentence.

The offender management accountability also includes managing different situations that may arise with a prisoner due to their individual needs, for example, monitoring and observing behaviours, and identifying and responding to prisoners who are 'at risk' of harm.

Under the administration accountability, corrections officers are required to contribute to the Department's record keeping (both paper-based and using the centralised electronic system) by writing reports and file notes.

The final two accountabilities – personal development and teamwork – relate to the personal and social characteristics of officers. A high standard of behaviour is required and corrections officers are required to act with integrity and professionalism. Officers are encouraged to develop and maintain their competencies. Teamwork is essential for maintaining a safe and secure working environment. Passing on information about a prisoner to other relevant staff, such as changes in a prisoner's behaviour, helps to identify changes in risk, thereby helping to maintain the safety and security of the prison unit.

3.2.4. New directions in the working environment

New Zealand corrections officers have been encouraged to use techniques known as 'active management' in their daily interactions with offenders since the late 1990s. These techniques include using every contact with offenders as an opportunity for positive influence, for example, through positive reinforcement. However, in recent years, this rehabilitative component of the role has received increased focus, and corrections officers have been trained in a framework called Right Track.

The Right Track framework supports and encourages corrections officers to have a greater role in promoting positive behaviour change. The framework consists of roles, processes, and practice strategies. As part of this framework, corrections officers are trained in techniques to develop effective working relationships with prisoners and to motivate and influence change. Right Track involves regular meetings between prison staff members in different roles, in which individual prisoner cases are discussed.

At the core of the Right Track framework is the Stages of Change model (Prochaska & DiClemente, 1983; Prochaska, DiClemente, & Norcross, 1992). Officers are trained to identify which stage a prisoner may be at in relation to a particular rehabilitative need (e.g. pre-contemplation), and in doing so identify the prisoner's 'readiness' for change. Officers then set appropriate goals and to use appropriate motivational tactics in their interactions with the prisoner to move them through the stages of change (e.g. from 'pre-contemplation' through to 'motivated to change').

The Right Track framework emphasises the importance of the 'right relationship' between the corrections officer and the prisoner. Although all three accountabilities of security, care and rehabilitation must be applied, an understanding is needed of when to focus on one accountability more than another. Desired staff behaviours include good communication, empathy, sound judgement, decision making, and integrity (O'Fallon, 2014).

Implementation of Right Track began in 2012, and as such, it is still in a relatively new stage. However, early evaluations of the framework indicate that staff are satisfactorily applying the principles and Right Track behaviours are embedding in practice (O'Fallon, 2014). In the longer term, it is expected that the effect of Right Track on prisoner behaviour will be measured, for example, incident frequency and compliance with offender plans (O'Fallon, 2014).

3.3. Predictors of corrections officer job performance

Lambert and Paoline (2008) suggested that corrections institutions usually succeed or fail due to the performance of their employees. However, there is comparatively less research about corrections officer performance than other aspects of correctional work, such as stress and role conflict. Given the importance of employees to achieving the aims of the organisation, job performance, and predictors of job performance for corrections officers is a topic for further research.

3.3.1. Specific antecedents to performance

Although psychologists have researched the predictors of performance for staff in similar occupations, including police officers, there is comparatively less research carried out in this area in relation to corrections officers. Additionally, most of the available literature on corrections officer performance comes from the United States. There is little recent psychological research in relation to the selection and performance of corrections officers.

Job-specific tests

One of the most comprehensive studies of selection procedures for corrections officers comes from a study carried out in Pennsylvania over a twenty-year period, during which time the selection procedures evolved (Sproule & Berkley, 2001). In the 1980s, general cognitive ability tests were replaced with job specific tests (written and oral). The job specific written test used multiple-choice questions to assess visual memory through photographs, associative memory (mug shots and names), and reading comprehension. The oral test assessed motivation, job interest, poise and self-confidence, communication skills and judgement.

A validation study was carried out in 1984 with a group of four hundred officers, of whom 94% were male (Sproule & Berkley, 2001). Performance ratings were obtained from personnel files after 6 months and 12 months on-the-job. Further criterion measures were a job-specific corrections officer rating form, as well as dismissals, sick leave, and turnover. The results showed that written test scores were significantly correlated with both the 12-month performance rating and the scores on the rating form. Oral test scores were significantly correlated with the overall rating on the form and the 12-month performance rating. The combined score on the written and oral tests (weighted at 50% each) was correlated with the overall rating on the rating form (r = .34). In summary, the results indicated that the job-related cognitive ability test predicted performance, as well as the structured oral assessment.

The assessment procedure was later refined to include video-based components depicting incidents and events in the correctional setting. The written test was updated to assess observation (using the video clip), following oral instructions (again using the video clip), understanding rules and regulations, and following written instructions. The oral test assessed ‘poise and confidence’, ‘oral communication skills’, and ‘judgement and problem-solving skills’. A writing exercise was also included, in which officers wrote a narrative based on an incident on videotape. This was designed to assess written communication.

A further validation study was carried out in 1992 to assess the utility of these selection procedures (Sproule & Berkley, 2001). The validation involved 500 corrections officers, and the sample was 92% male. The criteria included the academy test scores (a test completed after each week of the training programme), job performance ratings by supervisors (overall, and on nine factors), and a weighted combination of both test scores and job performance ratings. The results indicated that the written test was the best predictor of all three criteria. Subtest scores on the writing exercise and factors in the oral test were not found to be good predictors of ratings on similar aspects of job performance. For example, ‘judgement and problem-solving’ on the oral test did not predict these abilities on the job. In summary, the written test was an excellent predictor of performance. Adding the oral test and writing exercise only marginally increased the correlation.

Personality traits

Schuerger, Kochevar and Reinwald (1982) investigated personality in relation to corrections officer performance and how this relationship varied by gender. The Sixteen Personality Factor Questionnaire (16 PF) was administered to corrections officers working at a county holding facility in the United States. As the researchers predicted, the personality profile of corrections officers aligned with Holland’s Realistic Type, and the results revealed significant, positive similarities with other Realistic Type occupations, notably the police and mechanics. Interestingly, two peaks were identified in the factors relating to control – conscientious and controlled – and these were shared with police. Corrections officers were also slightly introverted and low in anxiety.

Performance ratings were obtained from agency files. The ratings had been filled out by first-line supervisors and covered quality of work, quantity of work, knowledge, need for supervision, attendance and punctuality, judgment, attitude, personality, personal appearance, and leadership ability. Most of the variance in these ratings was accounted for by a single global factor, although a lesser factor was found for attitude and attendance. For male corrections officers, ‘the ‘bright’ personality factor, ego strength, and self-sufficiency were related to the global performance factor. Conscientiousness and conservative were close to significance.

In summary, the highest performing male officers appeared on the personality profile as being bright, controlled, conservative, and self-sufficient. No significant relationships were found between personality variables and performance for females. The highest performing women in the sample tended to have low dominance and trust, although this may have been due to a preference by the male supervisors to rate women displaying these personality traits as higher (Schuerger et al., 1982).

The Pennsylvania study (Sproule & Berkley, 2001) also investigated the usefulness of personality measures in predicting success on the job for corrections officers. In 1984, the Jackson Personality Research Form (PRF-E) did not predict overall ratings on the performance rating form, but two items did predict ratings: endurance and understanding (in the negative direction). This indicated that corrections officers willing to work longer hours, and those who were less inquiring and analytical, tended to receive higher performance ratings.

In response to the widespread use of the Minnesota Multiphasic Personality Inventory (MMPI) for law enforcement officers (i.e. police), a study was undertaken to validate the use of both the MMPI and Inwald Personality Inventory (IPI) in sample of corrections officers (Shusman, Inwald, & Landa, 1984). As well as completing the personality assessments, applicants were interviewed during the application process by a psychologist who rated their suitability for the position, and background checks were also undertaken to screen out unsuitable applicants.

In the sample of 716 officers, 665 officers were retained and 51 were dismissed after a ten-month probationary period. The study investigated retention-termination rates, as well as job performance on three criteria (absences, lateness, and disciplinary interviews) in relation to the results of the MMPI, IPI, and both measures together. Using discriminant analysis, the IPI correctly classified 73% of the recruits as remaining on or removed from the job. The MMPI did not add to this. In relationship to job performance, the IPI correctly classified a greater number of officers to ‘positive’ and ‘negative’ job performance than the MMPI. In summary, the study shows that the MMPI, and IPI to a greater extent, can help in identifying officers who will have later performance difficulties.

Education

Findings are mixed regarding the relationship between education and performance of corrections officers (Philliber, 1987; Robinson, Porporino, & Simourd, 1997). Philliber (1987) in reviewing the literature on correction officers, noted that increased education for corrections officers is often suggested as a means of resolving all issues in the occupation. However, Philliber described a finding that education was associated with less job satisfaction and had no relationship to attitudes about the job.

Robinson, Porporino and Simourd (1997) investigated the assumption that postsecondary education improved job performance of corrections officers. It is often assumed that higher education will result in greater rehabilitation endorsement by corrections officers and capacity to deliver on the rehabilitative components of the role. The data came from 213 Canadian corrections officers using questionnaires, and the performance data came from the officers' latest performance appraisal.

The results indicated that correctional officers with higher levels of education were generally less satisfied with their jobs, and less likely to emphasise the custodial function of correctional work. There was no association between level of education and the other measured variables in the study, including performance, job motivation, and organisational commitment. In sum, although education predicted officer support of rehabilitation, it did not predict performance or other attitudes about the job. Age and gender were better predictors of desirable work outcomes than education. Robinson et al. (1997) noted that supporting rehabilitation in principle, as did the more highly educated officers, did not necessarily mean they were more willing to engage in rehabilitation activities with offenders. The authors suggested that selecting officers with the right disposition towards correctional work is a better approach than selecting for higher education.

Integrity

Integrity is considered an important requirement for corrections officers, given the nature of the work, the need to be prosocial role-models, and the risk of potential manipulation by offenders. The Reid Test is an overt measure of integrity that was used for screening corrections officers in the United States. It measures *punitiveness*, which contends that honest people maintain high personal standards and expect consequences for those who do not. It also measures *projectiveness,* which contends that individuals who are honest believe that most others are honest too.

In a corrections officer sample, no significant association was found between the Reid Test Overall score and subscores on each of the performance appraisal categories, or incidence of disciplinary action (Van Hein, Kramer, & Hein, 2007). The authors acknowledged restrictions in the data might have led to the absence of validity in the sample, for example, range restriction, and a tendency for managers to give employees 'passing scores' when providing performance ratings.

CHAPTER 4: THE PRESENT STUDY AND RESEARCH AIMS

The present study investigates general reasoning ability and job performance in a New Zealand corrections officer sample. There are two sets of research aims. The first set of research aims is concerned with the use of the psychometric assessments in the local occupational setting. The study involves the use of the General Reasoning Test (GRT2). This psychometric tool was developed in the United Kingdom, and while the test manual reports its psychometric properties in various settings, its psychometric properties (e.g. construct validity and reliability) in the local setting are unknown. The study also uses a custom-built performance appraisal tool, and the factors of performance as measured by this tool in the Corrections setting will be investigated. The second set of research aims involves testing the hypothesised relationships between the dimensions of reasoning ability, and between reasoning abilities and job performance. As well as investigating whether reasoning abilities predict job performance in this setting, the study seeks to provide criterion-related validity for the GRT2 assessment. The study also investigates the effect of participant characteristics (gender).

4.1. Evaluating tests for local use

Assessments with known psychometric properties are intended to provide a standardised metric for comparative purposes, for example, comparing individuals in pool of job applicants. Despite the intention of many test developers to develop tests for use in more than one country, there is debate in the literature about the cross-cultural transportability of assessments.

It is well recognised that there are difficulties associated with importing an assessment for use in a population different to that in which the assessment was developed (Dana, 1993; Hambleton, Merenda, & Spielberger, 2005; Samuda & Wolfgang, 1985; Suzuki & Ponterotto, 2008). For example, tests developed in one culture, may not be equally valid in another culture, as the test takers are likely to have different values and knowledge than those implicitly assumed by the test (Greenfield, 1997). The implication is that the test might not measure what it was intended to measure (e.g. ability), but instead measure respondents' (deficit) knowledge about another culture's cultural conventions (Greenfield, 1997).

Many test publishers are based in the United States and United Kingdom, and the tests are imported for use in other countries. Most of the commercially available psychometric tests in New Zealand were developed by international test publishers, and this issue has been recognised by New Zealand researchers (cf. Rodriguez, Treacy, Sowerby, & Murphy, 1998). Although local (New Zealand) norms are usually available for referencing respondents' scores, as the tests were designed and initially validated in other countries, items within the test and the test itself may not have the same validity in the local setting.

Importing and using a test in New Zealand therefore requires careful consideration as to whether it is appropriate to use the test in the local setting. The setting does not only include the country (i.e. the appropriateness of using a test designed in the UK in New Zealand), but also the appropriateness of using a test in particular occupational setting.

Assessments from other countries can be adopted or adapted for use in a new country (van de Vijver, 2002). Various strategies have been proposed to assist the process of importing and adapting tests for local use. Greenfield (1997) emphasised assessing the degree and type of difference between the test's origin culture and the culture in which it is has been imported to, and where there is a large degree of difference, considering the use of completely different assessment methods.

Test adaptation and modification is another strategy, and researchers have proposed guidelines to assist test importation and translation (cf. Hambleton, 1996, 2001; Hambleton & De Jong, 2003). Guidelines focus on processes such as test development, reliability assessment, validity assessment and norming, establishing item equivalence, establishing score equivalence, test administration and test-taking guidelines, and the methodologies associated with these test adaptation processes.

Attempts to adapt a test for a local setting must involve a review of the adapted test and statistical analysis to ensure that the modified test maintains its integrity. For example, Rodriguez et al. (1998) investigated the applicability of Australian adaptations of the Wechsler Intelligence Scale for Children (3rd edition) and the Stanford-Binet Intelligence Scale (4th edition) in a sample of Dunedin children. The researchers found that the test means of the Dunedin sample were comparable to the US norms, providing some initial reassurance about the use of the modified tests in this country.

The General Reasoning Test

The General Reasoning Test (GRT2) is a commercially available psychometric test developed in the United Kingdom by the test publisher Psytech. Although it has been used in various occupational settings in New Zealand, and New Zealand norms are available, it is newly introduced to the New Zealand Corrections setting. To establish its applicability to the local occupational setting, at a minimum the test needs to be re-validated using a sample from the population in which the test is intended to be used (i.e. corrections officers working in New Zealand). As its psychometric properties are unknown in the new setting, the study will examine the factor structure and construct validity when the test is used in the local sample and re-estimate its reliability.

The General Reasoning Test (GRT2) is designed to assess three forms or subtypes of general reasoning ability (verbal, numerical, abstract reasoning). The assessment is comprised of three separate subscales, with the items in each subscale designed to assess the same facet of reasoning ability. It is expected that the underlying structure (latent dimensions) in the GRT2 data will also resemble the three-scale structure in the local setting. To test this assumption, factor analysis will be conducted using item-level data from all three subscales of the GRT2, in an attempt to reproduce a three-dimensional structure. As the original factor loadings are not available in the technical manual for the GRT2, exploratory rather than confirmatory factor analysis will be used (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Thompson, 2004).

Professionally constructed assessments of cognitive ability tend to have excellent psychometric properties, including good levels of reliability (Ones et al., 2012). The technical manual for the GRT2 reported that the test had good levels of internal consistency reliability in a range of occupational samples, with most Cronbach's α coefficients above .8 (Psytech International Ltd., n.d). It is expected that the GRT2 will also show good internal consistency reliability in the present sample. In addition, the study will examine the item difficulty and item discrimination statistics for each subscale.

- *Assumption 1*: The GRT2 is expected to reveal a latent structure comprising three dimensions representing verbal, numerical, and abstract reasoning, on the Corrections sample.
- *Assumption 2*: The GRT2 subscales will have good levels of internal consistency reliability as evidenced by the Cronbach's α, on the Corrections sample.

Job performance appraisal

It is well recognised that job performance is a multi-faceted construct (Borman & Motowidlo, 1993, 1997; Sackett, 2002; Schmitt, 2014). Several models have been proposed for the dimensionality of job performance, for example, task and contextual performance (Borman & Motowidlo, 1993) and adaptive performance (Pulakos et al., 2000). Searches through the literature did not find evidence of any published research into the facets of job performance for corrections officers working in New Zealand.

The study uses a custom-designed job performance appraisal tool. The tool is concise but it does permit assessment of sub-constructs of performance. Information on underlying factor structure or other evidence for construct validity is not available in existing publications for the tool in its current form. An exploratory aim is to gather evidence for factor structure over the item set. Factor analysis will be used to investigate dimensions of performance in corrections officer data. This may determine whether performance as assessed by this tool is best explained as a single, global factor, or as an array of multiple facets. The conceptual nature of any facets will be examined. The reliability of the performance measure will also be analysed.

- *Research aim*: To explore the dimensions in the measured performance data for corrections officers working in New Zealand.

4.2. Substantive research aims

Dimensions of cognitive ability

Numerous studies have found that scores on psychometric assessments of cognitive ability tend to correlate positively with each other (cf. Jensen, 1986; Ree & Earles, 1991; Spearman, 1904). Many psychologists have accepted this 'positive manifold' as representing a common or general factor underlying cognitive ability (Carroll, 1993; Gottfredson, 1998; Jensen, 1987). However, alternative explanations have also been proposed (van der Maas et al., 2006), and the phenomenon may be the result of more than one factor or process (Horn & Blankson, 2012).

The GRT2 is designed to measure three facets of general reasoning ability, rather than narrow facets or aptitudes (Psytech International Ltd., n.d). Therefore, the subscales are expected to tap into a sizable 'common' component, and are expected to correlate significantly with each other. However, the correlations will not be so high as to indicate that the subscales are measuring the same construct.

Conceptualisations of intelligence tend to partition fluid (*Gf*) and crystallised (*Gc*) intelligence as separate forms or subtypes of intelligence (Horn & Cattell, 1966; Nisbett et al., 2012; Schneider & McGrew, 2012). However, *Gf* and *Gc* also correlate (Lubinski, 2004), suggesting a commonality or association between the different forms of ability. Whereas fluid intelligence is thought to represent the most general form of reasoning ability (e.g. the ability to solve abstract and novel problems), crystallised intelligence involves the accumulation of knowledge and the use of this knowledge to solve problems. Other researchers have equated *Gf* with general mental ability or *g* (Gustafsson, 1984; Valentin Kvist & Gustafsson, 2008).

It is possible to view the GRT2 subscales in terms of fluid and crystallised intelligence. Theoretically, the abstract reasoning subscale aligns to the concept of fluid intelligence. Although the verbal and numerical reasoning subscales are designed to assess facets of general reasoning ability, these scales require learned knowledge such as the meanings of words and rules for solving mathematical problems. The numerical and verbal reasoning subscales are conceptually closer to crystallised intelligence.

It is predicted that abstract reasoning will have the highest correlations with the other subscales, whereas verbal and numerical reasoning will correlate to a lesser extent. This prediction is based on the greater conceptual similarity between fluid intelligence and forms of crystallised intelligence, in comparison to two distinct forms of crystallised intelligence.

- *Hypothesis 1*: Verbal, numerical, and abstract reasoning ability are expected to be positively related.
- *Hypothesis 2*: Within the range of correlations among verbal, numerical, and abstract reasoning scores, abstract reasoning will be more strongly correlated with the other subscales than the correlation between numerical and verbal reasoning.

Cognitive ability and job performance

Research has consistently identified cognitive ability as a predictor of performance across multiple job settings (cf. McHenry et al., 1990; Olea & Ree, 1994; Ree et al., 1994; Schmidt & Hunter, 2004), with the prediction being stronger in jobs with more complexity (Schmidt & Hunter, 1998). As identified in the literature review, the corrections officer role is associated with unique complexities, including balancing the different accountabilities of the role (e.g. security, care, rehabilitation). Based on the evidence that cognitive ability predicts performance in a range of occupational settings, it is hypothesised that reasoning abilities will also predict overall job performance for corrections officers.

Better predictions about the relationship between cognitive ability and outcomes such as job performance can be made when looking at specific facets of the variables. Using the findings from the factor analysis about the dimensions of corrections officer performance, relationships will be explored between each of the reasoning ability variables and the performance dimensions.

- *Hypothesis 3*: All reasoning ability components are expected to be associated with job performance.
- *Research aim*: To explore the relationships between reasoning abilities and the facets of performance for corrections officers.

Participant characteristics, ability, and performance

The research findings regarding gender differences in reasoning abilities are inconsistent. Some studies have indicated that males are stronger in numerical reasoning and females are stronger in verbal reasoning, whereas meta-analyses have identified that these differences are negligible (Hyde et al., 1990; Hyde & Linn, 1988). Therefore, it is predicted that any differences between male and female participants on the GRT2 subscales will be non-significant.

Other research of interest regarding the relationship of cognitive ability to gender has found that overall reasoning ability does not vary by gender, but it may mask differences between specific abilities (Johnson & Bouchard, 2007), and that males tend to have greater variability in scores than females, particularly in the extreme high end of the scale (Hedges & Nowell, 1995). These studies suggest that there may be gender differences in terms of cognitive ability structure, as well as variability of scores. The study will investigate the reasoning ability dimensions separately for male and female participants, including the correlations between the subscales. The study will also look at the relationship between reasoning abilities and overall job performance by gender.

- *Hypothesis 4*: There will be no gender difference in components of reasoning ability.
- *Research aim*: To explore the pattern of relationships among reasoning abilities separately for male and female participants, including correlations among verbal, numerical, and abstract reasoning.
- *Research aim*: To explore the relationships between reasoning abilities and overall job performance separately for male and female participants.

CHAPTER 5: METHOD

5.1. Permissions

An application was made to the Department of Corrections Research and Evaluation Committee to re-analyse data that were previously collected for an internal project. The internal project involved the validation of assessments used during the selection of corrections officers. Permission was granted by the Research and Evaluation Committee to re-analyse part of this data set consisting of ability test scores and corresponding performance ratings. An application was also made to the Massey University Human Ethics Committee and resulted in the ethical approval for this present research project (MUHEC Application 13/78). The data set was provided to the researcher in electronic spreadsheet format, fully anonymised, without names or other information that could identify individuals.

5.2. Participants

The population was defined as all corrections officers working in New Zealand. There were seventeen prison sites at the time of the study. For practical reasons, the study used data from a sample of this population. The data came from assessment of a sample of 88 corrections officers working in New Zealand prisons at seven prison sites. The sampling procedure and representativeness of the sample are discussed in more depth in the Discussion chapter (section 7.3).

Prison managers were requested to nominate corrections officers to participate in the study, representing a range of performance levels, and reflecting the gender and ethnic composition of prison staff. All participants had the job title (rank) of 'corrections officer'. With the exception of gender, no other demographic and background variables were available because this information was not requested from the test provider during the original project. The majority of the participants were male (64%), which reflects the high proportion of male corrections officers working in New Zealand prisons.

Ability scores for all three subscales of the GRT2 (verbal, numerical, abstract reasoning) were available for 88 corrections officers (56 male officers, 32 female officers). Corresponding performance data were available for 78 of these participants (47 male officers, 31 female officers). Item-level data (correct/incorrect responses for individual GRT2 items) were available for the 88 participants who completed the GRT2.

Participants were provided with a candidate information sheet and release form. The information sheet explained that the purpose of the study was to trial a new cognitive ability test for use in selecting corrections officers and to determine the attributes required to be an effective corrections officer. The information sheet and release form are included in Appendix B.

5.3. Data collection

The following section details the psychometric and assessment tools used to collect the data, the psychometric properties of the tools, and the procedure that was followed during data collection.

5.3.1. The General Reasoning Test (GRT2)

The General Reasoning Test (GRT2) is a cognitive ability test developed by Psychometrics Ltd and published by Psytech International. OPRA Consulting Group is the main distributor for New Zealand. OPRA supplied the computer-based, online version of the GRT2 used to collect the data for the Department's original study.

The General Reasoning Test (GRT2) is designed to assess general reasoning ability in adults (over 15 years of age). It is used in job selection to differentiate between applicants for jobs that require a general (or 'non-graduate') level of cognitive ability. Examples of jobs requiring a general level of ability include junior sales and administration roles, clerical, and technical roles. The GRT2 is the alternate version of the Graduate Reasoning Test (GRT1), which differs from the GRT2 in that it is designed to distinguish between applicants for jobs that require a higher than average level of reasoning ability, such as applicants for managerial and professional roles.

It is important that tests are designed to assess ability in the population that they will be used. If a test designed for one group is used to assess a group that is different (e.g. in terms of occupational group), there is a high likelihood of obtaining a floor or ceiling effect, making it difficult to distinguish between those individuals in the group being assessed. If the Graduate Reasoning Test (GRT1) was used to assess individuals in jobs requiring a general level of ability, it is likely that scores would cluster in the lower range, and individual differences between test-takers would not be as noticeable. Similarly, if the General Reasoning Test (GRT2) was given to a professional occupational group, it is likely that scores would cluster around the higher end, making it difficult to identify the higher ability applicants within the group. During the development of the GRT1 and GRT2, items were trialled in groups with different educational levels and occupations, and items were identified for each version of the test.

The GRT2 consists of three subtests that measure constructs described as three different dimensions of general reasoning ability: verbal reasoning, numerical reasoning, and abstract reasoning. The subscales are identified in the GRT2 technical manual as VR2, NR2, and AR2, respectively. For consistency, these labels will be used in the present study when referring to the subscales (see next subsection). Together the three subscales assess both crystallised and fluid forms of intelligence. Verbal reasoning and numerical reasoning are affected by an individual's previous education and experience, and therefore, the NR2 and VR2 subscales are expected to assess the crystallised form of intelligence. Out of the three subscales, the abstract reasoning test (AR2) is the least likely to be influenced by previous experience and education, and aligns more closely with the construct of fluid intelligence.

When completing the GRT2, respondents complete the subtests in successive order, and each subtest is separately timed. The verbal reasoning subtest consists of 35 items and respondents have 8 minutes in which to answer; the numerical reasoning subtest consists of 25 items to be answered in 10 minutes; and the abstract reasoning subtest consists of 25 items to be completed in 10 minutes. At the start of each subtest, respondents are advised of the number of items and how long they have to complete the subtest. The test instructions advise respondents to work as quickly and accurately as possible.

5.3.2. Verbal, numerical, and abstract reasoning subtests

The verbal reasoning subtest (VR2), according to the GRT2 technical manual, is designed to assess the ability to understand the meaning of words, to classify words into categories, to understand the relationships between pairs of words, and to draw conclusions and inferences from words (Psytech International Ltd., n.d). Each item consists of a question and six possible responses, and respondents select one answer from the options provided. An example of a verbal reasoning item is shown below.[2]

Which of the following is the odd one out?

1. Farmer
2. Sailor
3. Accountant

[2] Test items provided as examples are drawn from a practice assessment (OPRA Consulting Group, 2009) and not the actual GRT2. The example items are representative of the types of questions asked.

4. Employee
5. Police officer
6. Shop assistant

(OPRA Consulting Group, 2009)

The numerical reasoning subtest (NR2) is designed to assess the ability to categorise numbers, understand relationships between pairs of numbers, and to understand and form number sequences (Psytech International Ltd., n.d). The item format of the NR2 is the same as the other GRT2 subtests, with each item having six possible response options. An example of a numerical reasoning item is shown below.

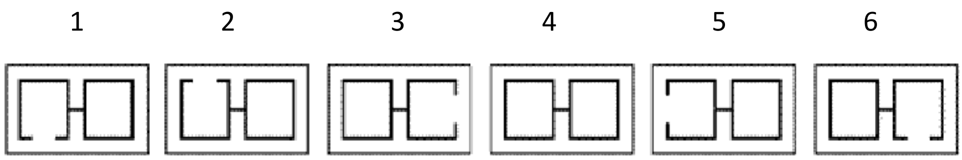
Travelling (constantly) at 30 km/hr how long would it take to travel 45 kilometres?

1. 20 mins
2. 30 mins
3. 50 mins
4. 1 hour
5. 1 hour 30 mins
6. 2 hours

(OPRA Consulting Group, 2009)

According to the test manual, the abstract reasoning subtest (AR2) is designed to assess the ability to understand abstract logical problems. Items assess the ability to understand logical relationships between abstract or geometric patterns, the ability to classify patterns, and to form sequences of patterns. Each item has six response options; an example is shown below.

Which of the following is the odd one out?



The initial results of the GRT2 are raw scores for each subscale. The raw score is the number of items the respondent answered correctly in the subscale. The test is norm-referenced and raw scores are not usually reported. Instead, the raw scores are compared to a norm (or reference) group of respondents who have already completed the test. The raw scores are then transformed into 'normed' stanine scores (on a scale of one to nine) that allow comparison between the respondent and the reference group (Lyman, 1998). This is to determine where the respondent 'sits' in terms of percentile rank compared to the rest of the norming sample.

It is important that the norm chosen for a particular assessment situation is based on a norming sample as similar as possible to the respondent or applicant group (Groth-Marnat, 2009). This allows meaningful and appropriate comparisons to take place. An example of a general norm is a country norm (e.g. New Zealand adults), whereas examples of more specific norms are age, gender, and occupational group norms. The general New Zealand norm group for the GRT2 consists of approximately 5,000 New Zealand respondents. In selection, some organisations prefer to use an occupational norm group to ensure a closer match between the applicant and reference group. At the time of the study, a specific occupational norm group for corrections officers did not exist as the GRT2 had only recently been introduced to the organisation. Raw scores were used in the statistical analysis for this present study. This approach was considered appropriate because the focus of the research was not to compare individual participants with each other, but rather to explore the aggregate data at group level.

5.3.3. Psychometric properties of the GRT2

Reliability

According to the GRT2 technical manual, the test has good levels of test-retest reliability (temporal stability). The test-retest reliability was assessed by administering the GRT2 with the same group of 54 college students on two separate occasions, two weeks apart. The test-retest coefficients were .81, .84, and .78 for the verbal, numerical, and abstract reasoning subscales, respectively (Psytech International Ltd., n.d). The high correlation coefficients suggested that the respondents' scores were relatively stable on the two occasions they completed the test, and that there was relatively low error in the test score measurement. This result was consistent with the conceptualisation of the cognitive ability construct as relatively stable across time.

The GRT2 technical manual reported high levels of reliability on the basis of internal consistency as well. Cronbach's α coefficients were reported for a number of occupational and demographic groups, and ranged from .78 to .86 for the verbal reasoning subscale, .82 to .89 for the numerical reasoning subscale, and .78 to .85 for the abstract reasoning subscale. For the New Zealand Adults reference group (n = 5,183), the α coefficients were .83, .88, and .82 for the verbal, numerical, and abstract reasoning subscales respectively (Psytech International Ltd., n.d). The high α levels indicated that individuals scoring highly on one test item tended to score highly on other test items within that subscale. This was evidence that each GRT2 subscale measured the same attribute or dimension and was not overly influenced by measurement error. The internal consistency of the GRT2 when used with the corrections officer sample was assessed as part of this study. Details of this analysis are described in the Results chapter (subsection 6.3.2).

Validity

The GRT2 technical manual provided evidence of the test's construct validity and that each of the GRT2 subscales significantly correlated with the other two subscales. The reported correlations were .62 between the verbal and numerical reasoning subscales, .63 between the abstract and numerical reasoning subscales, and .59 between the abstract and verbal reasoning subscales (Psytech International Ltd., n.d). The degree of the correlations suggested that although the subscales appeared to measure different facets of reasoning ability, the correlations were not so high as to suggest they were measuring the same construct. This was taken as evidence of both the convergent and the discriminant validity of the GRT2 subscales.

Further evidence of the construct validity of the GRT2 came from a study that correlated the subscales of the GRT2 with the subscales of an assessment called the TTB2 (Technical Test Battery). The TTB2 subscales were visual acuity, mechanical reasoning, and spatial reasoning. Statistically significant correlations were found between each subscale of the GRT2 and TTB2; however, these were modest in size (Psytech International Ltd., n.d). The technical manual suggested that the modest correlations were due to the TTB2 assessing relatively specific technical aptitudes, whereas the GRT2 assessed broader reasoning abilities.

There is evidence that the scores on the GRT2 vary as expected with the education levels of respondents. The GRT2 technical manual provided mean raw scores by educational level, showing that respondents who attained higher levels of education, obtained higher scores on the GRT2. For example, respondents who had not completed a secondary education had mean scores of 15.0, 9.1, and 12.6 on the verbal, numerical, and abstract reasoning subscales, respectively. Respondents with a tertiary education had higher mean scores of 20.9, 13.3, and 15.3 for the verbal, numerical, and abstract reasoning subscales (Psytech International Ltd., n.d). This was strong evidence of the construct validity of the GRT2.

The construct validity of the GRT2 in the corrections officer sample was assessed as part of this study. Details of this analysis are provided in the Results chapter.

5.3.4. Job performance appraisal questionnaire

A 23-item performance appraisal questionnaire was used to obtain managerial ratings of participants' job performance. The performance measure was developed for use in the Department's original study. The items on the measure were written in view of a classic 'domain' for content validity, i.e. as attributes identified to be important to the corrections officer role. The appraisal tool has three parts.

The first part of the measure consists of one item only, asking for an overall rating of the corrections officer's performance using a 5-point item-level Likert scale (unsatisfactory; needs improvement; meets standard; slightly above standard; well above standard).

The second part of the measure features three items asking managers to rate how frequently the corrections officer went above the requirements of their role, demonstrated a desire to learn, and assisted others without personal benefit. Each item uses a 5-point scale of frequency (almost never; occasionally; sometimes; often; almost always).

The third part of the measure includes 19 items relating to perceived skill in a range of different areas considered important for the corrections officer role. These included oral and written communication, numerical skill, learning ability, recovery from setbacks, success with Māori offenders, trustworthiness, situational awareness, self-awareness, interpersonal skill, ability to stay calm, and team loyalty. Each item is rated on a 5-point scale of frequency (almost never; occasionally; sometimes; often; almost always). The performance measure is included in Appendix C.

The reliability of the job performance measure was checked as part of the study (Results chapter, subsection 6.4.2).

5.3.5. Data collection procedure

Participants completed the assessment on-site at their place of work (prison site) during scheduled supervised sessions, using the online, computer-based version of the GRT2. The assessment process was managed by recruitment advisors trained in GRT2 administration. Following the standardised verbal instructions, participants followed on-screen instructions to complete the assessment. In total, participants spent 1 hour and 30 minutes completing assessments. This total time included the GRT2 as well as an assessment of emotional intelligence and a written work sample test[3]. A maximum of 28 minutes was allowed for completion of the GRT2 via the computer-based programme, excluding the initial administration instructions. Participants were not provided with their assessment results following completion of the GRT2; however, they were able to request this information at a later date.

Managerial ratings of corrections officer performance were obtained via the online performance appraisal questionnaire (see Appendix C). The ratings were provided by principal corrections officers with line management responsibility for the officers in the sample.

[3] The original validation study included the GRT2, an emotional intelligence assessment, and a written work sample test. Only the GRT2 component was analysed in this present study.

CHAPTER 6: RESULTS

6.1. Description of variables

Raw scores on the GRT2 verbal, numerical, and abstract reasoning subscales, GRT2 item responses, and performance ratings were received in electronic spreadsheet format from the test supplier. The data were transferred into SPSS software (version 22) for analysis.

Table A1 (in Appendix D) offers a comprehensive listing of variables available for quantitative analysis, including hypothesis testing. It provides SPSS variable names, along with a simplified description, the possible range of permitted values, and the Stevens level of measurement ('scale type'; Stevens, 1946).

Four of the variables described in Table A1 were created during the statistical analysis: *Overall_GRT2* and the three performance factor scores, *Affective*, *Cognitive*, and *Citizenship*. Descriptions of how these variables were created are provided in subsections 6.2.1 and 6.4.1 of this Results chapter.

Cognitive ability data were available for 88 participants, and performance data were available for 78 of these participants. Missing data were dealt with pair-wise to maximise the data available for analysis.

Strategy of data analysis

The first stage in the data analysis was initial analysis to calculate descriptive statistics for the cognitive ability and job performance data. Normal distribution is an assumption for some statistical tests, and the distributions of the variables were checked as a preliminary step for the analysis. The data were also checked for any anomalies, issues, or missing data before proceeding to the next stage of the analysis.

The second stage was to evaluate the local adequacy of the assessment tools: (a) the GRT2, and (b) the job performance tool. For the GRT2, this stage involved examining the latent structure of the cognitive ability data using factor analysis *(Assumption 1),* re-estimating the reliability of each subscale *(Assumption 2)*, and carrying out item-level analyses by calculating item difficulty and item discrimination statistics. For the job performance appraisal questionnaire, this stage involved examining the latent structure of the job performance data using factor analysis, deriving factor score variables for use in the next stage of the analysis, and estimating the reliability of the job performance appraisal tool.

The final stage in the data analysis was to test the substantive hypotheses and examine the hypothesised relationships between variables, beginning with the correlations between the subscales of the GRT2 *(Hypotheses 1 and 2)*. Correlational analyses were also conducted to explore the relationship between reasoning ability and job performance *(Hypothesis 3)*. This analysis also had the purpose of estimating the criterion-related validity of the GRT2. Lastly, the effect of gender was explored, as this was the only participant characteristic information was available for in the data set. This involved comparisons of the mean reasoning ability subscale scores for male and female participants *(Hypothesis 4)*, calculating the pattern of correlations among the reasoning ability subscales separately for male and female participants, and lastly, exploring the hypothesised relationships between reasoning ability and job performance separately for male and female participants.

6.2. Initial data analyses

6.2.1. Cognitive ability: Descriptive statistics

The mean scores and standard deviations were verbal reasoning, *M* = 19.67 (*SD* = 5.10), numerical reasoning, *M* = 12.15 (*SD* = 4.93), and abstract reasoning, *M* = 14.62 (*SD* = 5.08). Table 1 provides the descriptive statistics regarding each subscale variable.

As shown in Table 1, the three variables representing the GRT2 subscales each had an adequate range. The skewness and kurtosis statistics for each subscale were negligible, particularly for the verbal and abstract reasoning subscales, indicating that the distributions were approximately normal in shape. The exception was the numerical reasoning subscale, which had a skewness statistic of more than twice the standard error of skewness. This could be considered a slight positive skew, and indicated that scores were clustered in the lower rather than higher range of the distribution. Figures A1 to A3 in the Appendix illustrate the distribution of scores on each GRT2 subscale.

Table 1

Range, Mean, Standard Deviation, and Distribution of GRT2 Subscale Scores

	Min.	Max.	Mean	*SD*	Skewness	*SE* of skewness	Kurtosis	*SE* of kurtosis
Verbal reasoning	4	32	19.67	5.10	-.22	.26	.26	.51
Numerical reasoning	3	25	12.15	4.93	.56	.26	-.22	.51
Abstract reasoning	0	24	14.62	5.08	-.28	.26	-.28	.51

An overall GRT2 variable was derived from the data set for use in the empirical analysis. The raw scores for each subscale were converted to *z*-scores and then summed to create a composite score with equal weightings for each reasoning dimension. The composite score was transformed to a standardised variable and named *Overall_GRT2*.

6.2.2. Job performance data: Descriptive statistics

Conservatively, the item-level Likert scales used in the tool are ordinal (they supply rank order), but they are often treated as interval scales because it is assumed that there is equidistantiality of any gap between adjacent scale points. When Likert scale data meet normal distribution and equidistantiality assumptions, it is admissible to treat the data as interval scale data. Therefore, Table A2 (in Appendix) provides the mean and standard deviation for the performance measure items, along with the median and range statistics, which are considered more suitable for ordinal scale data.

The median item ratings indicated that employees tended to be rated favourably on the performance measure. The first performance item, which asked managers to rate the employee's overall performance, had a median rating of 3 ('meets standards'). All other items had median ratings of either 3 or 4, with the majority of the items having a median rating of 4 and corresponding to 'often' in relation to how often the employee performed that skill or behaviour. The ratings on the performance items had adequate ranges, with nearly all items having the full range of ratings from 1 to 5.

The distribution statistics were studied to determine whether the data were normally distributed. Although most items had a very slight negative skew, reflecting a slight tendency for participants to receive favourable performance ratings, the skewness statistic was greater than two times the standard error of skewness (indicating significance) for only four items (items 14, 15, 22, and 23). Performance item 14 measuring 'trustworthiness, person of their word' had a greater negative skew than the other items, indicating that most ratings were in the higher range for this item. Item 14 was the only item with a significant positive kurtosis value. With the exception of the four items identified, the remaining items can be described as approximately normally distributed.

Although the data appear to be approximately normally distributed for most performance items, the assumption of normality is not required for correlational analyses when one of the variables is ordinal data. In generic terms, both non-parametric and parametric tests could be used in analyses, taking into account that the data can be treated as either ordinal or interval scale data. Additionally, with the exception of performance item 1 (overall performance rating), all other items were transformed into performance factor scores through factor analysis of the performance data. The composite performance variables (factor scores) were then used in the empirical analysis. The development of the composite performance variables is detailed in subsection 6.4.1.

Performance item 1 (overall performance) was treated as a separate variable in the empirical analyses, and Table 2 provides more information about the distribution of ratings on this item. Although the overall distribution of ratings was approximately normal, there was an absence of ratings at the extreme lower end of the scale (i.e. ratings of 1).

Table 2

Frequency of Ratings for Performance Item 1 (Overall Performance)

Rating	1	2	3	4	5	Total
Frequency	0	17	27	25	9	78
Percent	0	21.8	34.6	32.1	11.5	100
Cumulative percent	0	21.8	56.4	88.5	100	

6.3. Evaluating local adequacy of ability test

6.3.1. Factor structure of GRT2

One of the research aims was to explore the structure of cognitive ability by identifying a latent structure using factor analysis, expecting a replication of three dimensions *(Assumption 1)*. Principal axis factoring (PAF) was used to detect mutually interconnected subsets of items (Kline, 1994; Thompson, 2004).

An initial inspection of the correlation matrix revealed some inter-item correlations of $r = .3$ or greater; however, out of the 3570 item pairs, only 148 had correlations of this magnitude. None of the item pairs had correlation coefficients greater than .8, which indicated that multicollinearity was not an issue. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .09. This statistic was considerably less than .6, the minimum recommended KMO statistic in order for the data to factor well. An examination of the anti-image matrix diagonal revealed that the KMO statistics were less than .5 for all variables, meaning that no items stood out for removal from the analysis and that removing items would not have improved the overall KMO. However, the Bartlett's Test of Sphericity was significant (X^2 = 5136.89 (3570), $p < .001$). The final commonalities for nearly all items were low (< .5) and single items did not stand out for removal. Although the data were not ideal in terms of factorability, the factor analysis was continued to explore for any emergent patterns of dimensionality.

Different strategies can be used to determine how many factors to retain in a factor analysis, each with strengths and weaknesses (cf. Hayton, Allen, & Scarpello, 2004; Horn, 1965b; Zwick & Velicer, 1986). Retaining and interpreting only the factors with eigenvalues greater than 1 is a common preference (Zwick & Velicer, 1986), but in large matrices it can result in overestimates of the number of factors (Kline, 1994). The scree test involves examining the scree plot to identify the number of factors after which the curve appears to 'level off' (Kline, 2005). Ideally, different strategies can be used in the hope they corroborate each other (Thompson, 2004).

The results showed that 29 factors had eigenvalues greater than 1, with 'levelling off' occurring in the scree plot around the sixth factor. It was not possible to reach a simple solution with this number of factors. Therefore, only the first three factors, although explaining a relatively low percentage of overall variance (21.2%), were examined for interpretability. Table 3 shows the eigenvalues and variance explained by the first three factors. The decision to focus on three factors was based on the premise that the GRT2 measures three dimensions of ability (verbal, numerical, and abstract reasoning). The first factor had an initial eigenvalue of 9.88, explaining 11.63% of the total variance. The second and third factors were smaller, with initial eigenvalues of 4.37 and 3.77, and explaining 5.14% and 4.44% of the variance in GRT2 scores, respectively.

Table 3

Eigenvalues and Variance Explained Before and After Rotation for GRT2 Factors 1 to 3

	Initial eigenvalues			Extraction			Rotation[a]
Factor	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total
1	9.88	11.63	11.63	9.14	10.75	10.75	8.32
2	4.37	5.14	16.76	3.64	4.28	15.04	3.55
3	3.77	4.44	21.20	3.14	3.69	18.73	5.35

Note.
[a] As the factors were correlated, the rotated sums of squares loadings could not be added to obtain a total variance.

The factors were rotated using the Direct Oblimin method in SPSS. It was assumed that the dimensions of cognitive ability were correlated; therefore, oblique factor rotation was considered the most appropriate rotation method (Kline, 1994; Thompson, 2004).

Table A3 (in Appendix) presents the factor loadings in terms of the unique contribution of each item to the rotated factors (pattern matrix), and Table A4 (in Appendix) displays the factor loadings as simple correlations (structure matrix). Only factor loadings of .3 or greater are shown. When examining the factor loadings in Table A3, the first factor appeared to be a more general factor, with items loading highly on the factor mostly coming from the abstract and numerical reasoning subscales, and a smaller number of items from the verbal reasoning subscale. The second factor appeared to represent numerical reasoning, with numerical reasoning items loading more highly on the factor than items from the other two subscales. The third factor had the highest loadings from verbal reasoning items.

The content pattern of the factors remained when using the more stringent cut-off of .5 or greater for the factor loadings. As highlighted in Table A3, Factor 1 had eleven items with factor loadings of .5 or greater, and of these, four were abstract reasoning items and seven were numerical reasoning items. All four items with factor loadings of .5 or greater on Factor 2 were numerical items, and the five items with factor loadings of .5 or greater on Factor 3 were verbal reasoning items.

The factor analysis provided limited support for an underlying three-dimensional structure, with the first three factors explaining only approximately one-fifth of the total variance in GRT2 scores. Although the data did not factor well, the content or subject-matter patterns of the first three factors were roughly in the direction expected. This was more evident for Factors 2 and 3, which appeared to represent more specific abilities (numerical and verbal reasoning), whereas Factor 1 appeared to represent a broader reasoning ability.

6.3.2. Reliability

Based on original British research by Psytech, it was predicted that the GRT2 would have good psychometric properties, including reliability *(Assumption 2)*. An analysis of internal consistency reliability was undertaken using Corrections data.

Reliability of GRT2 subscales

For each subscale of the test, reliability was re-estimated over the Corrections data set. Cronbach's α was calculated as .76 for the verbal reasoning subscale, .82 for the numerical reasoning subscale, and .81 for the abstract reasoning subscale (N = 88). These α levels suggested an acceptable level of internal consistency for the verbal reasoning subscale, and an good level for the numerical and abstract reasoning subscales.

Apart from contributing to classical item analysis, item-total score correlations can be viewed as supporting a check on test reliability (see Tables A5 to A7 in the Appendix). For the verbal reasoning subscale, although several items had corrected item-total correlations close to zero, the removal of these items would not have considerably improved Cronbach's α. Similarly, for numerical reasoning and abstract reasoning, removing item(s) would have only resulted in marginal changes to Cronbach's α.

6.3.3. Item analysis

There are two major options in analysing the functionality of items on psychometric tests – Classical Test Theory (CTT) and Item Response Theory (IRT). Given the small and limited data set, a simplified CTT approach remained realistic for analysing items in the GRT2. Two item parameters, item difficulty and item discrimination, were estimated in the Corrections data.

Item difficulty

The first item parameter estimated over Corrections data was item difficulty (Kline, 1999; Nunnally, 1978). The analysis explored the range of difficulty of items in the GRT2. On an ability aptitude test, the item difficulty index is simply the proportion of respondents in a sample who answered the item correctly; often this is referred to as the *p* value (Kline, 1999). The index can range from 0 to 1 for each item, with 0 indicating that no respondents answered the item correctly, and 1 indicating that all respondents answered correctly. For this data set, the item difficulty index also corresponded to the mean score for the item.

Figures 1 to 3 illustrate the distribution of the items by difficulty for each subscale. For verbal reasoning, items ranged from .11 through to .93 on the difficulty index, indicating that the items varied widely in how easily they were answered. Most items on the verbal reasoning subscale were between .5 and 1 on the difficulty index, indicating a high proportion of items that could be answered by most participants. The numerical reasoning subscale also had a wide range of items in term of difficulty, with items ranging between .11 and .81 on the index and most items sitting in the range of .3 to .8. Out of the three subscales, the numerical reasoning subscale had the greatest proportion of difficult items (index < .5). The abstract reasoning subscale had items ranging from .32 to .92 on the difficulty index, and unlike the other two subscales, lacked items at the extreme end of the difficulty index (< .3). Out of the three subscales, the abstract reasoning scale had the highest proportion of items that could be answered by most participants (i.e. difficulty index > .5).

The mean difficulty indices for each subscale were .56 (verbal reasoning), .49 (numerical reasoning), and .59 (abstract reasoning). These statistics suggested that the numerical subscale had the most difficult items, on average, followed by the verbal and abstract subscales, which were relatively similar in terms of average item difficulty.

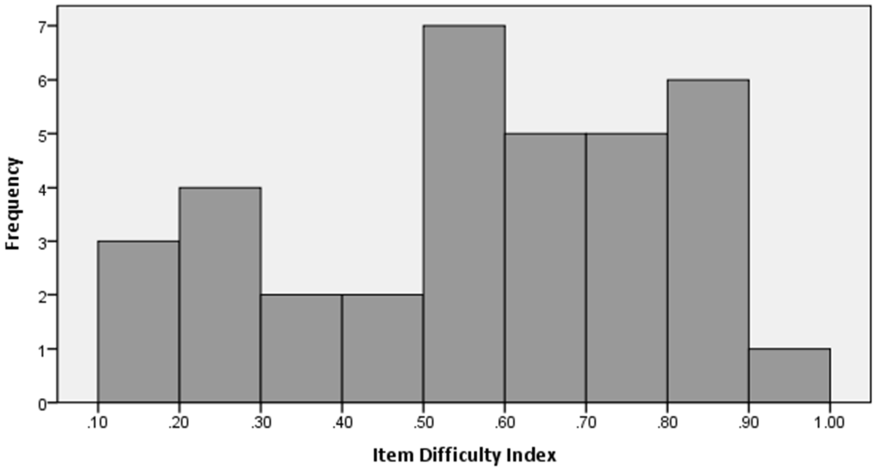


Figure 1. Difficulty of items in the verbal reasoning subscale
Note. N of items = 35, *M* = .56, *SD* = .23.

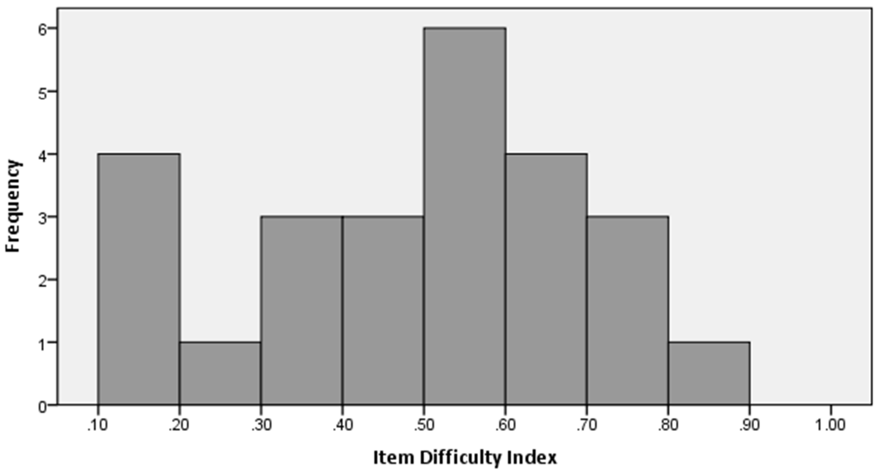


Figure 2. Difficulty of items in the numerical reasoning subscale
Note. N of items = 25, *M* = .49, *SD* = .21.

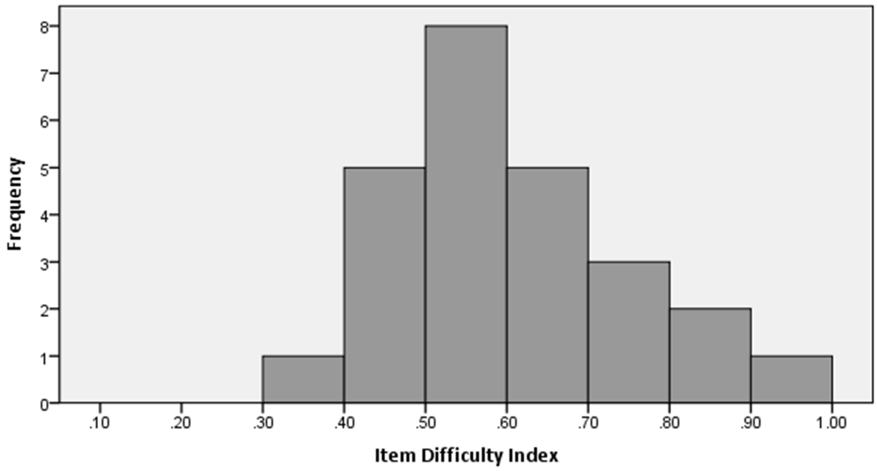


Figure 3. Difficulty of items in the abstract reasoning subscale
Note. *N* of items = 25, *M* = .59, *SD* = .15.

Item discrimination

A second parameter estimated for each GRT2 item was discrimination (Kline, 1999; Nunnally, 1978). Item discrimination refers to how performance on individual test items relates to performance on the total scale (i.e. the correlation between the item and the overall score) (Kline, 1999). The parameter indicates how well each item discriminates between those who scored well on the scale overall and those who did not.

There are different methods of calculating the item discrimination parameter. Indices to correlate individual items with the total score in item analysis include the Pearson product-moment correlation, the point-biserial correlation and phi coefficient (Kline, 1999). The point-biserial correlation is the most appropriate index to use when one of the variables is dichotomous (Kline, 1999; Nunnally, 1978). In this study, the dichotomous variable is the incorrect/correct responses coded as 0/1. It was appropriate to use a corrected version of the point-biserial correlation to remove the item from the total test score, as not to inflate the correlation coefficient. This corrected point-biserial statistic was the same as the corrected item-total correlation.

Tables A8 to A10 (in Appendix) provide the item discrimination index (both the item-total and corrected item-total correlations) calculated over Corrections data, alongside the item difficulty index (proportion of correct responses for the item). Item discrimination and item difficulty can be understood in the context of each other, as items with a difficulty index of .5 usually have better discrimination power.

In the verbal reasoning subscale, two items had weak negative corrected item-total correlations (items 3 and 7). The negative correlation suggested that individuals scoring highly overall on the scale did not tend to answer the item correctly. This could be an indicator of low quality for these items. For item 3, the undesirable item discrimination index (-.05) may have been due to a ceiling effect, as the item difficulty index was high (.86) suggesting that nearly all respondents were able to answer the item. For item 7, the undesirable item discrimination index (-.02) may have been due to a floor effect. The item had a very low item difficulty index (.11), meaning that few respondents were able to answer the item. Verbal items 8, 13, 20, and 27 also had low discrimination indices (< .1), signalling possible issues with the quality of these items.

All items in the numerical reasoning subscale had acceptable discriminating power, with all items having an item discrimination index of greater than .1. In the abstract reasoning subscale, only item 4 had a discrimination index of less than .1. Interestingly, the difficulty index was .55, which was in the range of difficulty when discrimination power is predicted to be at its highest (Kline, 2005). This could also suggest an issue with the item's contribution to the overall measurement of a construct, as although approximately half of the participants could answer it correctly, successfully answering the item had no relationship with the overall subscale score. It could also indicate an issue with the way a test item affects overall test validity (construct-related or even content-related validity); for example, that perhaps the item is more sensitive to an extraneous construct other than abstract reasoning.

6.4. Evaluating the performance appraisal tool

6.4.1. Dimensionality of job performance

One of the research aims was to explore the factor structure (performance dimensions) in the corrections officer sample. The tool for performance appraisal tool was available to the researcher but no information was accessed about underlying factors. Principal axis factoring (PAF) was used to clarify and understand latent dimensions in the item-item correlation matrix (Kline, 1994; Thompson, 2004).

An initial inspection of the item-item correlation matrix (Table A11 in Appendix) revealed that all correlation coefficients were at or above .3. The overall performance rating (PF1) had significant correlations with all other items ($p < .001$). Performance item 9 ('takes the time to understand the needs of Māori') and item 10 ('proactively integrates their understanding of Māori needs when relating to Māori') were very highly correlated ($r = .92$), indicating a potential multicollinearity issue.

Performance items 2 to 23, but excluding item 10, were entered into the factor analysis. Given that purpose of the factor analysis was to explore the facets of job performance, item 1 was excluded, as it was the overall performance rating. Item 10 was also excluded due to its high correlation with item 9. This meant that 21 items were entered into the factor analysis. Performance data were available for 78 participants; however, a greater number (more than 100) would have been ideal to meet the minimum of five cases per variable that is recommended for factor analysis. Bartlett's Test of Sphericity was significant (X^2 = 1767.54 (210), p < .001), and the Kaiser-Meyer-Olkin (KMO) statistic was .93, indicating very high interrelatedness over all variables and suitability for factor analysis.

As shown in Table 4, three factors had initial eigenvalues greater than 1, and these three factors cumulatively explained 76.73% of the variance in performance. The factor analysis revealed a strong first factor with an initial eigenvalue of 13.75 and explaining 65.49% of the variance in performance, and two substantially weaker factors with eigenvalues of 1.36 and 1.00, explaining only 6.46% and 4.78% of variance in performance, respectively.

Table 4

Eigenvalues and Variance Explained Before and After Rotation for Performance Factors 1 to 3

	Initial eigenvalues			Extraction			Rotation		
Factor	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	13.75	65.49	65.49	13.49	64.25	64.25	6.48	30.85	30.85
2	1.36	6.46	71.95	1.04	4.94	69.20	5.04	24.00	54.85
3	1.00	4.78	76.73	0.73	3.46	72.66	3.74	17.81	72.66

Note.
The table shows only factors with eigenvalues greater than 1.

The factors were rotated using Varimax rotation in SPSS. An orthogonal rotation method was used because there were no advance assumptions about the performance factors being correlated (Kline, 1994; Thompson, 2004). As shown in Table 4, the rotated factors accounted for approximately 30.85%, 24% and 17.81% of variance in performance, respectively. The rotated factor matrix is provided in Table A12 (in Appendix). Only factor loadings of .3 or greater are shown.

When examining the content of the items that loaded highly on the first factor, it was apparent that this first performance factor had an emotional or affective component, comprising of understanding and managing one's own emotions, and to a smaller extent, understanding others' emotions. The item with the highest loading (.82) on Factor 1 was 'can effectively manage their own emotions, both positive and negative'. As highlighted in Table A12, other items with high loadings (.6 or greater) on this factor had content associated with emotional and self-management domains, including with being in tune with one's own emotions, expressing feelings appropriately, remaining calm under pressure, being trustworthy, loyal to the team, understanding others' emotions, and recovering from setbacks.

Based on the highest-loading items, the second performance factor appeared to relate to cognitively-based tasks on the job. The item with the highest loading (.74) on Factor 2 was 'picks up new learning quickly and without the need for repeated explanation'. Other performance items with high loadings (.6 or greater) on the factor were also cognitively-oriented, including thinking and planning, exploring alternatives before taking action, clear communication with numbers and data, and situational awareness. The item concerning 'going above and beyond the requirements of their role' also appeared to cluster with this factor, indicating that respondents (i.e. managers) may have perceived above-role requirements in more terms of the cognitively-oriented domain of performance rather than the affective domain as represented by the first factor.

The third performance factor appeared to have an interpersonal focus and relate to building effective relationships at work. The item with the highest loading (.72) on Factor 3 was 'builds positive relationships, regardless of nationality, age and gender'. Other items with high loadings related to understanding needs of Māori, considering others' feelings when decision making, and voluntarily assisting others.

Three job performance variables were created from the factor scores. These were provisionally labelled *affective*, *cognitive*, and *citizenship*. The labels were chosen based on the nature and formulation of items that loaded highly on each factor. This was particularly the case for the first two factors, and the finding that the first factor appeared to represent performance in terms of an affective (or emotional) domain, and the second factor appeared to represent performance in terms of the cognitive domain. The third factor was reminiscent of organisational citizenship behaviour at the individual level and concepts such as altruism and courtesy towards others. Accordingly, Factor 3 was labelled *citizenship*.

Table 5 displays distribution statistics for each of the performance factor variables. The three performance variables each had a slightly negative skew. However, both the skewness and kurtosis statistics were in the range for assuming a normal distribution and suitability for parametric tests.

Table 5

Distribution Statistics for Performance Factors 1 to 3

	Min.	Max.	*M*	*SD*	Skewness	*SE* of skewness	Kurtosis	*SE* of kurtosis
Performance factor 1: Affective	-2.52	1.80	.00	.94	-.52	.27	-.06	.54
Performance factor 2: Cognitive	-2.46	2.93	.00	.93	-.08	.27	.57	.54
Performance factor 3: Citizenship	-2.48	2.32	.00	.90	-.13	.27	.26	.54

Relationship of performance dimensions to overall performance

The job performance appraisal tool included an overall rating of job performance. The relationships of this 'global performance' item to the performance factors (derived from ratings on the remainder of the job performance items) were examined to gain an insight into how managers interpreted the global performance rating, and whether ratings on this item corresponded to all or some of the performance factors.

An inspection of the item-item correlation matrix (Table A11 in Appendix) of the performance appraisal tool revealed the overall performance rating had high correlations with both cognitively- and affectively-oriented performance items. The overall performance rating (PF1) correlated most highly with PF13 'explores alternative options prior to taking action' (r = .75), closely followed by PF18 'can tune into what make people tick and connect with others' (r = .72). The association of the overall performance rating item with both the cognitive and emotion-oriented performance domains suggests that the item was well-understood by respondents (managers), and that their ratings on this item reflected the main facets of performance in this context.

Pearson's correlation coefficients *(r*) and Spearman's rho (ρ) were calculated to estimate the association between overall performance and the performance factors. When exploring bivariate relationships, Spearman's rho is the appropriate correlation coefficient to compute when at least one variable is ordinal. Although conservatively viewed as ordinal data, the preliminary analyses (subsection 6.2.2) found that the performance measure variables (including the overall performance rating) were approximately normally distributed and met equidistantiality assumptions. The performance factor score variables (Table 5) were also approximately normally distributed and could be managed as interval-level scales. This meant that it was also admissible to use Pearson's correlation coefficients.

Significant positive correlations were found between the overall job performance rating (PF1) and each of the three performance factors. As shown in Table 6, the overall performance rating had the strongest relationship with the cognitive performance domain ($r = .60$, $p < .001$), followed by the affective domain ($r = .48$, $p < .001$), and lastly the citizenship domain ($r = .36$, $p = .001$). Therefore, the overall job performance rating was likely to reflect consideration of all the main facets of performance, with a slight bias towards cognitively-oriented performance.

Table 6

Correlations between Overall Job Performance and Performance Factors

		Overall job performance rating
Performance factor 1: Affective	r	$.48^{**}$
	ρ	$.46^{**}$
Performance factor 2: Cognitive	r	$.60^{**}$
	ρ	$.62^{**}$
Performance factor 3: Citizenship	r	$.36^{*}$
	ρ	$.44^{**}$

Note.
* Correlation is significant at the $p < .01$ level (2-tailed).
** Correlation is significant at the $p < .001$ level (2-tailed).

Principal Components Analysis

To check the possibility that the results of the factor analysis were influenced by the choice of particular procedure within factor analytic methods (principal axis factoring), an alternative analysis was conducted using principal components analysis (PCA). The results of the PCA were similar to those found with principal axis factoring. Three principal components were extracted with initial eigenvalues greater than 1. There was a strong first component (explaining 65.49% of the variance in performance) and two substantially weaker components making smaller contributions to the cumulative 76.73% explained by the three components. The three components had similar subject-matter patterns as the three factors identified through the principal axis factoring method (i.e. covering affective, cognitive, and citizenship-oriented performance domains).

6.4.2. Reliability of the job performance measure

A Cronbach's α was calculated using 22 items on the performance appraisal tool (items 2 to 23). The overall performance rating (PF1) was excluded as it was treated independently of the other performance measure items during the empirical analyses. The α coefficient was calculated as .97, indicating a near-ideal level of reliability in terms of internal consistency. Furthermore, as shown in Table A13 (in Appendix), all items had acceptable item-total correlations. Removal of any individual item(s) would not have improved the α.

However, caution was required when interpreting this α coefficient. Increasing the number of items inflates the α (Cortina, 1993), and the statistic is also influenced by the dimensionality of the data. As revealed in the factor analysis, the performance measure was not unidimensional but consisted of three separate performance dimensions.

6.5. Substantive results and hypothesis testing

6.5.1. Relationships among GRT2 reasoning dimensions

Hypothesis 1 predicted that relationships would be found between verbal, numerical, and abstract reasoning. Table 7 details the correlations between scores on each subscale.

Table 7

Correlations between GRT2 Subscales

	Verbal reasoning	Numerical reasoning	Abstract reasoning
Verbal reasoning	—	.43	.51
Numerical reasoning		—	.44
Abstract reasoning			—

Note.
All correlations are significant at the $p < .001$ level (2-tailed).

Significant positive correlations were found between verbal and numerical reasoning ($r = .43$, $p < .001$), verbal and abstract reasoning ($r = .51$, $p < .001$), and numerical and abstract reasoning ($r = .44$, $p < .001$). The results supported *Hypothesis 1* and *Hypothesis 2*, which predicted that the relationships between abstract reasoning and each of numerical and verbal reasoning would be higher than the relationship between numerical and verbal reasoning.

These differences in strength of relationship were minor, however. The mean correlation between the three subscales was r = .46.

6.5.2. Cognitive ability and job performance

Correlations between cognitive ability and overall performance

To test for relationships between cognitive ability and overall job performance *(Hypothesis 3)*, bivariate correlations (both Pearson's correlation coefficients and Spearman's rho) were calculated between each of the cognitive ability variables (verbal, numerical, abstract, and overall reasoning) and the overall performance rating (PF1). As discussed in subsection 6.4.1, although the performance variables were conservatively treated as ordinal data, it was also admissible to compute Pearson's correlation coefficients. Performance data were missing for 10 participants; therefore, data from 78 out of the 88 participants in the study were used in the correlational analyses.

No relationships were found between any of the cognitive ability variables and overall job performance. As shown in Table 8, the correlation coefficients were all close to zero and statistically non-significant.

Correlations between cognitive ability and performance dimensions

Correlations were also calculated between the cognitive ability variables and the dimensions of performance (*affective*, *cognitive*, and *citizenship*) using the three composite performance variables derived from the factor analysis (subsection 6.4.1). The results are shown in Table 8. A weak correlation (ρ = .23, p = .043) was found between numerical reasoning and the first performance factor (affective). An inspection of the scatterplot does not suggest a relationship between these variables, however. Given the small sample size, it is possible that this finding arises from a Type I error. No other statistically significant relationships were found between the cognitive ability variables and any of the performance dimensions.

Table 8

Correlations between GRT2 (Subscales and Overall) and Job Performance (Overall and Dimensions)

		Verbal reasoning	Numerical reasoning	Abstract reasoning	Overall reasoning
Overall job performance	*r*	.00	.05	.05	.04
	ρ	.01	.09	.02	.06
Performance factor 1: Affective	*r*	-.09	.16	.03	.04
	ρ	-.02	.23*	.05	.10
Performance factor 2: Cognitive	*r*	.10	.06	.17	.14
	ρ	.06	.10	.11	.11
Performance factor 3: Citizenship	*r*	.06	-.12	-.04	-.04
	ρ	.09	-.09	-.08	-.02

Note.
* Correlation is significant at the $p < .05$ level.

6.5.3. Participant characteristics, ability, and job performance

Gender and reasoning ability

An analysis to determine whether there were any gender differences in subscales of the GRT2 *(Hypothesis 4)* was conducted. An initial inspection of the mean subscale scores by gender (Table 9) revealed that the mean score on the verbal reasoning subscale was higher for female participants, whereas the mean scores on the numerical and abstract reasoning subscales were higher for male participants. The data came from 56 male and 32 female participants.

To check for significant differences in the mean subscale scores for male and female participants, two-tailed *t*-tests were conducted. The Levene's test indicated equality of variances across the groups. The results of the *t*-tests indicated that males scored significantly lower than females on the verbal reasoning subscale, $t(86) = -3.07$, $p = .003$, but scored significantly higher than females on the numerical reasoning subscale, $t(86) = 2.19$, $p = .031$. Although male participants scored slightly higher on abstract reasoning than female participants, this difference was not statistically significant, $t(86) = .39$, $p = .697$. *Hypothesis 4* that there would be no gender difference in components of reasoning ability was supported for abstract reasoning, but not for verbal or numerical reasoning.

Table 9

Mean GRT2 Subscale Scores for Male and Female Participants

	Males		Females	
	M	*SD*	*M*	*SD*
Verbal reasoning	18.46	5.02	21.78	4.60
Numerical reasoning	13.00	5.26	10.66	3.92
Abstract reasoning	14.79	5.42	14.34	4.48

Note.
n = 56 (males), n = 32 (females).

Gender and reasoning ability structure

The pattern of relationships among dimensions of reasoning ability was explored separately for subgroups of male and female participants. Table 10 displays the correlations. For male participants, the relationship between numerical and verbal reasoning (r = .60, p < .001) was slightly stronger than the relationship between abstract and verbal reasoning (r = .58, p < .001), and stronger than the relationship between abstract and numerical reasoning (r = .42, p = .001). For female participants, the relationship between numerical and verbal reasoning (r = .41, p = .021) was weaker than the relationships between abstract reasoning and each of verbal reasoning (r = .50, p = .003) and numerical reasoning (r = .50, p = .003).

Hypothesis 1 predicted that the dimensions of cognitive ability would be positively correlated. As presented in subsection 6.5.1, the hypothesis was supported when looking at male and female participants together. The same hypothesis is corroborated when analysing data for male and female participants separately. However, *Hypothesis 2* that the relationships between abstract reasoning and each of verbal and numerical reasoning would be stronger than the relationship between verbal and numerical reasoning, was confirmed only for female but not male participants. For males, the highest correlation between the subscales was found between the more specific reasoning abilities.

When checking mean level of correlation (i.e. the mean of bivariate coefficients among subscale variables in the matrix), this mean was higher for males (r = .53) than for females (r = .47). Additionally, these mean subscale correlations for male and female participants separately were higher than the mean correlation among the subscales found when looking at male and female participants together (r = .46) in subsection 6.5.1.

Table 10

Correlations between GRT2 Subscales for Male and Female Participants

	Males			Females		
	Verbal reasoning	Numerical reasoning	Abstract reasoning	Verbal reasoning	Numerical reasoning	Abstract reasoning
Verbal reasoning	_	.60***	.58***	_	.41*	.50**
Numerical reasoning		_	.42**		_	.50**
Abstract reasoning			_			_

Note.
* Correlation is significant at the $p < .05$ level (2-tailed).
** Correlation is significant at the $p < .01$ level (2-tailed).
*** Correlation is significant at the $p < .001$ level (2-tailed).

Gender and job performance

Possible gender disparity in performance ratings was also examined. The mean scores for male and female participants on each performance measure item are displayed in Table A14 (in Appendix).

Male and female participants received similar ratings on overall performance (PF1). The mean rating for males on this item was 3.32 (*SD* = .91) and for females it was 3.35 (*SD* = 1.02). Females received slightly higher ratings than males for 21 out of the 23 performance measure items. Two items were exceptions: PF14 ‘is a trustworthy individual; is a person of their word’, and PF20 ‘can effectively manage their emotions, both positive and negative’. Male participants received slightly higher ratings than female participants on these two items.

To determine whether gender differences in the performance ratings were statistically significant, two-tailed *t*-tests were conducted. Performance data were available for 47 males and 31 females. Levene’s Test for Equality of Variances was not significant and equal variances were assumed. The *t*-tests were not statistically significant for any of the performance measure items. This meant that although there were slight gender differences in mean scores on the performance item ratings, and a trend for females to receive slightly higher ratings, the apparent differences cannot be statistically established, and may constitute "noise". The results of the *t*-tests are also included in Table A14 (in Appendix).

Due to the small group numbers when the sample was broken down by gender, a factor analysis of the item-item correlation space for performance items was not carried out for male and female participants separately. Consequently, composite performance variables (from factor analysis) were not calculated for male and female participants separately.

Cognitive ability and overall job performance by gender

An additional analysis was carried out to determine if the hypothesised empirical relationships between cognitive ability and job performance *(Hypothesis 3)* would be found when examining the data for male and female participants separately.

Both Spearman's rho and Pearson's correlation coefficients were calculated between the dimensions of cognitive ability and overall job performance (PF1) for gender subgroups. As shown in Table 11, all correlations were statistically non-significant. Therefore, *Hypothesis 3*, which predicted associations between cognitive ability and overall performance, was not supported when looking at male and female participants separately.

Relationships between cognitive ability and factorial composites of job performance within the female and male subgroups were not assessed because the factor analysis was not completed for male and female participants separately.

Table 11

Correlations between GRT2 Subscale Scores and Overall Job Performance for Male and Female Participants

		Males				Females			
		Verbal reasoning	Numerical reasoning	Abstract reasoning	Overall reasoning	Verbal reasoning	Numerical reasoning	Abstract reasoning	Overall reasoning
Overall job performance	*r*	.09	-.07	-.02	.00	-.14	.26	.17	.11
	ρ	.09	-.01	-.07	.04	-.11	.25	.18	.11

Note.
All correlations are statistically non-significant.
n = 47 (males), *n* = 31 (females).

CHAPTER 7: DISCUSSION

7.1. Interpreting findings from evaluation of the GRT2 and performance measure

(A) Local evaluation of the ability test

When assessing the cognitive ability of officers, the project relied on a well-established, well-researched, respectable psychometric instrument (cf. Section 5.3). The General Reasoning Test (GRT2), a product of one of the United Kingdom's most respected test development agencies, has been subjected to thorough reliability estimation, validation, and norming research (Psytech International Ltd., n.d). The GRT2 has been used in New Zealand for several years with diverse organisations and thousands of persons tested.

However, appropriateness for use in a new local environment cannot be taken for granted (Anderson & Herriot, 1997). Whenever a psychometric test is brought to a new context, it needs to go through an adaptation process in a broad sense (Cronbach & Drenth, 1973). Contexts for test use include not only countries other than the place of original test development (the issue of 'test importation'), but diverse groups, cultural and language groups, age cohorts, and organisations. A principle of adaptation is that invariance in a new environment of the properties of a tool originally demonstrated and published by the developers cannot be taken for granted. For instance, reliability and validity estimates may change when the test is used in a new setting with new groups of persons (Wolf & Little, 1996).

A full adaptation project for a British or American test is usually undertaken when the new context involves testees speaking languages other than English and with beliefs, attitudes, and value orientations sharply different to the original target population (Cuéllar & Paniagua, 2000; Dana, 1993, 2000; Samuda, 1998; Suzuki, Meller, & Ponterotto, 1996). However, the shared language and traditions of Britain and New Zealand do not mean that all circumstances of use and all characteristics of the testees are identical or even similar. Greenfield (1997) explicitly warned about tacit assumptions regarding automatic ‘portability’ of ability and achievement tests across various English-speaking countries. New Zealand use requires test adaptation, even if this is smaller-scale and meets fewer challenges (e.g. no full translation and re-translation required). OPRA Group has undertaken research in relation to the GRT2, such as re-validation and re-norming with multiple groups in New Zealand. However, the results of this work are not yet published in regular scholarly channels, are unlikely to become published in full form, and are unavailable to the thesis writer. Additionally, most adaptation work performed by OPRA Group relates to organisations different to Corrections.

Logistics of the project significantly constrained the scope of adaptation work that could be undertaken for the GRT2 (cf. Section 6.3). Some important aspects of validity, especially construct validity in terms of factorial structure, as well as reliability, were investigated. Apart from the overall practical constraints of a minor project, statistical analysis was further limited by the small size of the participant group and small *N* of the data set provided.

Assumption 1 was concerned with examining the latent dimensions of the GRT2 to determine whether a three-dimensional structure that was originally reported by developers (Psytech International Ltd., n.d) could be reproduced on the New Zealand Corrections data. This is viewed as an attempt to produce evidence for an aspect of construct-related validity (cf. Sections 4.1, 6.3). The original structure covered components of verbal, numerical, and abstract reasoning. An exploratory factor analysis was conducted using all 85 items from subscales of the GRT2 (Mulaik, 2010; Thompson, 2004; Walkey & Welch, 2010). *Assumption 1* was not supported. A total 29 factors with eigenvalues greater than 1 were generated. These factors, in spite of reaching the eigenvalue of 1, accounted for small fragments of the total variance of the 85-variable space (Fabrigar et al., 1999; Hayton et al., 2004; Horn, 1965b; Zwick & Velicer, 1986). The three strongest factors were examined for interpretability and for any emergent content patterns.

The first factor had highest loadings from abstract and numerical reasoning items, whereas the second and third factors appeared to represent numerical and verbal reasoning. Interestingly, a distinct factor representing abstract reasoning was not found. Aside from the clear limitations of the factor analysis arising from the small size of the participant group, there are several other possible explanations for these findings.

The abstract reasoning subscale was thought to align more closely to fluid intelligence, whereas the numerical and verbal reasoning subscales were thought to align more closely to crystallised intelligence. However, it is likely that the distinction between the reasoning dimensions, and the measurement of fluid and crystallised intelligence, is not as clear-cut as implied by the three-scaled structure of the GRT2. A number of numerical reasoning items clustered with abstract reasoning items in the first factor, whereas other numerical reasoning items clustered in the second factor. It is possible that the numerical reasoning items that clustered with abstract reasoning tapped into fluid intelligence to a greater extent than the numerical reasoning items that clustered with the second factor, which may have had greater demands in terms of crystallised knowledge.

An alternative explanation is that the GRT2 – as it functions with this group of testees – is closer to a test of 'general mental ability' rather than assessing distinct components of general reasoning abilities. The comparatively large first factor that included items from all three subscales favours this interpretation.

Further factor analysis with a considerably greater number of cases, as well as confirmatory rather than exploratory options in factoring, would be required in order to determine whether a one-factor, two- or three-factor model would be a better fit for the cognitive ability data.

Together, the first three factors accounted for only approximately one-fifth of the variance in the GRT2 scores. A major limitation of interpreting output from all factor analyses was the limited *N*, presumed to be under the recommended minimum in terms of variables (Gorsuch, 1983; Thompson, 2004). This may explain why a simpler factor structure could not be obtained in this sample, and why the original components (Psytech International Ltd., n.d) were not fully 'reproduced'. The cases-to-variables ratio was less than desirable. For the large number of variables (85 test items), data would have been required from a much larger number of participants (at least 425, using the rule of five cases per variable that is usually recommended for factor analysis). The item-level variables were dichotomously coded (correct versus incorrect), and therefore did not reach the Stevens level of a true interval scale, which may have been another reason why the data did not factor as hypothesised. In many cases, dichotomous variables are considered appropriate for a matrix of Pearson correlation coefficients to be input to factor analysis.

With a larger sample and/or representative sample, it is still marginally possible that the GRT2 data could anchor three clear dimensions of reasoning ability in a corrections environment. Further research would be needed to investigate this.

Assumption 2 focused on the properties of the GRT2 from a different angle, that of measurement error. It was predicted that the GRT2 subscales would attain good internal consistency reliability (i.e. Cronbach's $\alpha > .8$) on the Corrections sample. This assumption was supported for two of the subscales.

The α coefficient for the verbal subscale was .76. This was in the 'acceptable' range, although slightly lower than the levels reported in the GRT2 technical manual for the verbal subscale (.79 to .89). The α for the numerical subscale was .82, which is considered 'good', but still at the lower end of the range of .82 to .89 reported in the technical manual. The α of .81 for the abstract reasoning subscale was also 'good', and was consistent with the previously reported range from .78 through to .85. Although the α levels in this sample were slightly lower for two of the GRT2 subscales than those reported in the technical manual, the previous studies had larger sample sizes. The study demonstrated that the subscales have an acceptable to good level of reliability in this corrections officer sample, which adds to the existing evidence about reliability of GRT2 in different occupational groups.

Simple item analysis, along the lines of classical rather than IRT-based analysis, was conducted to look at indices of item difficulty and discrimination in the GRT2. Analysing the GRT2 in terms of these indices adds to evidence about the quality of the test construction. The results of the item difficulty analysis revealed that in each subscale there was variation in terms of the ease at which items could be answered. While unsurprising, the finding is contrary to the assumption of classical test theory that items have parallel properties. The mean difficulty indices of .56, .49, and .59 for the verbal, numerical, and abstract reasoning subscales, respectively, suggested that the numerical subscale had the most difficult items, on average, followed by the verbal and abstract subscales. In comparison to the other two subscales, the numerical reasoning subscale had a greater proportion of items at the extreme difficulty level. In contrast, the abstract reasoning subscale had fewer extremely difficult items, and the verbal reasoning subscale had the greatest range in terms of item difficulty. All three subscales had the peak of the difficulty index distribution between .5 and .6, which gave the distributions an approximately normal shape.

The finding that the numerical reasoning subscale was experienced as the most difficult overall can be interpreted in several ways. For example, it might be that the extremely difficult items relied on specialised knowledge (crystallised intelligence) rather than ability per se to a greater extent than items in the other two subscales. An example is rules for working out mathematical equations. A limitation of the study was that the ‘incorrect’ response data did not separate out items that were answered incorrectly and those that were not attempted at all (‘timed out’). This information would have allowed greater insight into how the items were experienced and/or attempted by participants (e.g. if there was a relationship between the difficulty index and those items that were not attempted at all).

Findings on the item discrimination parameter revealed how well items discriminated between respondents who scored high versus low overall on the subscale. In the verbal subscale, six items had a discrimination index of less than .1, which may signal an issue with the quality of these items. However, these items with a very low discrimination index also tended to have either a very high or very low difficulty index, suggesting that floor and ceiling effects were the likely reason for the diminished discriminating power. In the numerical subscale, all items had discrimination indices greater than .1, which can be taken as evidence of the quality of the items in the scale. In the abstract reasoning subscale, all but one item had a discrimination index of greater than .1. The item that was the exception had a difficulty index of .55, suggesting that the item's low discrimination power was likely due to an issue with the item's reliability, for example, that it may tap into construct other than abstract reasoning, rather than being too difficult or easy.

Although such a thesis project has no resources to undertake full validation for all three aspects of test validity, findings supplied some information about the construct validity of the GRT2, and confirmed comparability with British test development outcomes.

The mean subscale scores for the corrections officer sample were consistent with the mean subscale scores provided in the GRT2 technical manual (Psytech International Ltd., n.d) for similar occupational and educational groups. The corrections officer scores were similar to the scores reported in the manual for the occupational groupings of 'clerical' and 'manual – skilled', and slightly higher than the mean scores for the 'semi-skilled – unskilled' workers. The corrections officers scores were also consistent with the mean scores for workers with educational attainment at the level of 'industry/trade qualifications' and 'completed secondary school', and were slightly higher than the mean scores for the 'not completed secondary school' grouping. The comparability of mean scores for corrections officers with overseas findings on markedly different occupational samples supports the local validity of the GRT2.

The distributions of scores on each subscale were approximately normal. There was an adequate range. This evidence supports the appropriateness of the test within this group (i.e. it was not too easy or too difficult for most respondents). As most respondents scored around the middle sub-range, the use of the assessment is unlikely to screen out too many or too few respondents when used with officers. One point of caution, however, is that the study was carried out with officers already appointed, rather than a job applicant group, which would have a wider score range.

(B) Performance dimensions in correctional work

One of the research aims was to gain deeper insight into the adequacy and functionality of the performance appraisal tool. While considerable work on test development and validation may have occurred, the results of this work are not yet published through academic channels (cf. Section 5.3.4). The measurement properties of the instrument were not known to the researcher in advance.

The study could not aim to conduct a full-fledged validation or reliability study to complete tool development. There were no means, within logistical constraints, to validate for all aspects of test validity, or for norming, item analysis, and proposals for revisions. However, some aspects of validity and reliability were possible to check.

An objective was to explore the dimensionality of performance in the corrections context, as measured by this performance appraisal tool (cf. Section 4.1). The data (managerial ratings of officers' performance) were subjected to factor analysis using principal axis factoring. Three factors were extracted. The rotated factors explained a cumulative 73% of the total variance in multi-item space. The first factor was larger (explaining 31% of the variance), and the next two factors comparatively smaller (explaining 24% and 18%).

The first factor was labelled the *affective* performance domain to reflect the emotionally-oriented nature of the items that loaded on the factor. The items included managing emotions, being in tune with one's feelings, expressing feelings appropriately, and remaining calm during pressure. The second factor, which was smaller than the first, was labelled the *cognitive* domain of performance. Items loading highly on the factor included tasks typically associated with learning, planning, exploring alternatives, communicating with words, and communicating with numbers and data. The third factor, accounting for a smaller percentage of the variance, included building positive relationships with others, regardless of age, gender or ethnicity, understanding the needs of Māori, considering others' feelings when decision-making, and voluntarily assisting others. Based on item language, the items loading on the third factor appeared to share an altruistic/helping theme, reminiscent of organisational citizenship at the individual level (i.e. supporting colleagues in the organisation; Borman & Motowidlo, 1993; Organ, 1997). Accordingly, the third factor was labelled *citizenship*.

Of particular note, was that the first performance dimension not only had an emotional orientation, but that the items with the highest loadings were related to understanding and managing one's own emotions. Items with the highest unique contributions to this factor were: (1) can effectively manage own emotions, both positive and negative; (2) is in tune with his/her own feelings, moods, and emotions; (3) expresses feelings to the right degree, to the right people, and at the right time; and (4) can remain calm and focused under pressure.

It could be argued, based again on interpretation of item language, that the first performance factor had more of an internal focus and was more about self-awareness and self-management, whereas the third performance factor had more of an external focus, with its content around understanding others and building effective relationships. The distinction is not entirely clear, however. For example, the item ‘can tune in to what makes people tick and connect with others’ loaded more highly onto the first factor than the other two factors, most likely due to the strong affective tones of the item.

The results of the factor analysis indicated that self-management and self-awareness attributes and behaviours represented a large component of performance in the corrections context. Self-awareness is important when working with offenders. It has a safety component (both psychological and physical). For example, ensuring that the right balance is maintained in staff-offender relationships and being aware when boundaries are ‘blurred’. In terms of physical safety, staying calm is an important part of defusing threatening incidents. Self-awareness is also important to maintain high standards and role-model prosocial behaviours.

The existence of an affective domain of performance, and the finding that it is a strong component of officers’ job performance, is unsurprising given the realities of the job and the correctional working environment. This is particularly relevant, for example, in day-to-day contact with prisoners, particularly as one aspect of the role is to motivate behaviour change. The finding is also consistent with the idea that the corrections officer role has a human-service or therapeutic direction, as opposed to only a custodial focus.

Factor-analytic results also provided useful insight into how performance is perceived of in a corrections environment. The dimensions of performance (affective, cognitive, citizenship), provide an alternative conceptualisation to what is offered by current performance models in management and organisational psychology literature. In this occupational setting, performance may not be best understood in terms of task versus contextual behaviour, but in terms of a performance model that is specific and customised to the occupational setting.

The loading of individual items on particular performance factors may also reveal how performance is perceived by managers in general, in a correctional environment. The item ‘going above and beyond the requirements of the role’ clustered with the second performance factor (cognitive). This suggests that in driving ratings by managers, behaviours exceeding the role minimum may be perceived more in terms of cognitively-oriented tasks, rather than those with an affective or organisational citizenship focus.

The item ‘pays attention to the environment and situational cues’ loaded onto both the first factor (affective) and the second factor (cognitive). Although the loading on the second factor was slightly higher, this finding does indicate that situational awareness in the corrections context is not purely a cognitive function. There is an affective component as well. This is likely a result of the unique environment, and the need to be cognizant of and attuned to emotional cues, for example, being alert to the prisoners' behaviours and changes in their behaviour.

It appears that the first item, the overall performance rating, was understood by managers in terms of the main facets of performance in the Corrections environment, as it correlated with all three factors of performance. There may have been a slight bias towards cognitively-based performance when managers made the overall performance ratings. A slightly higher correlation was found between overall performance and the cognitive performance factor, followed by the affective performance factor, and lastly the citizenship factor.

The item 'is a trustworthy individual, is a person of their word' clustered with the other self-management items on the first performance factor. This item stood out when examining the distribution of ratings on individual performance items. It was the only item in the performance measure that was significantly skewed, and ratings were clearly in the favourable direction. One reason may have been the specific type of work environment, and that considerable emphasis is placed on integrity and trustworthy behaviour. Managers may also feel a greater personality responsibility with regards to their officers' performance on this item.

The above interpretations are relativised by obvious caveats. Single-item 'measurement' goes against the principles of psychometrics. As a concise, economical tool, the performance appraisal tool encapsulates each notion in a single item, but how these items function in a respondent group remains empirically unclear. Words and phrases in items may be interpreted in dissimilar ways by different respondents, triggering ratings that become less comparable across testees.

The latent structure of the performance questionnaire, as revealed by the current analysis, did not match its 'face structure'. As presented, the questionnaire has a tripartite structure: (1) the overall performance rating, (2) ratings of three items relating to extra-role behaviours (going above requirements, investing in personal development, helping others), and (3) the remaining 19 items, which cover a range of skills and behaviours, including cognitive and affective. As revealed by multivariate analysis, the three items focusing on extra-role behaviours did not cohere as a distinct dimension. These insights may contribute to a refinement of the performance tool.

The reliability of the performance measure was estimated. The Cronbach's α of .97 indicated a high level of internal consistency, but caution is needed around this statistic. First, it can be questioned whether this is a unidimensional instrument where an assumption for internal consistency across all items is reasonable. Factor results imply three domains of performance, and therefore, a multi-dimensional, multi-construct tool. Further, as Cronbach's α increases with the number of items, the given item number may produce an inflated α coefficient.

7.2. Interpreting findings from hypothesis testing

(A) Structural components of cognitive ability

The study used data from a new occupational setting to examine the components of reasoning ability and especially any patterns of relationship among these components. In support of *Hypothesis 1*, significant correlations were found among scores on the verbal, numerical, and abstract reasoning subscales of the GRT2. This indicates that the subscales were not unrelated, and they may be seen as tapping into a common source underlying these components. However, the correlations were not so high as to suggest that the three subscales were covering the same construct (Sternberg, 1983).

The correlations found in the present study were $r = .43$ (verbal and numerical reasoning), $r = .51$ (verbal and abstract reasoning), and $r = .44$ (numerical and abstract reasoning). These correlations were lower than the subscale correlations reported in the GRT2 technical manual, which were $r = .62$ (verbal and numerical reasoning), $r = .59$ (verbal and abstract reasoning), and $r = .63$ (numerical and abstract reasoning).

One explanation for the lower correlations is that the dimensions of cognitive ability and the relationships between any two of them have manifested differently in this local occupational group, compared to a broader sample reported in the technical manual. This would suggest that the 'common' or 'general' component in the reasoning abilities was lesser in this group, than in other occupational samples. Another possibility is the influence of the very small N in the present study ($N = 88$, compared to $N = 5,183$).

Although the magnitude of coefficients was lower than previously reported, correlations found in the present study shed some light on the construct validity of the GRT2. All three subscales measured forms of general reasoning ability and the positive correlation found between the subscales was theoretically expected. However, the fact that the correlations were not near-perfect reflects the conceptual dissimilarity of the subscales as measures of different facets of general reasoning ability (Sternberg, Lautrey, & Lubart, 2003).

The finding that the subscales positively correlated is consistent with previous research across many projects, dating back more than a century, that has shown similar results (i.e. that scores on different cognitive tests tend to correlate and reveal a ‘positive manifold’; Sternberg, 1994). The finding is consistent with the notion of a ‘general mental ability’ construct.

The results of the study also support *Hypothesis 2*, which predicted that abstract reasoning would have the highest correlations with the other subscales. The finding is consistent with the evidence that fluid and crystallised intelligence are related forms of intelligence. The correlations between abstract reasoning (fluid intelligence) and each of numerical reasoning (crystallised intelligence) and verbal reasoning (crystallised intelligence) were higher than the correlations between numerical and verbal reasoning (two forms of crystallised intelligence; Cattell, 1987; Horn, 1965a; Horn & Cattell, 1966). The differences were very small, however.

(B) Cognitive ability and job performance

Hypothesis 3 that reasoning abilities would be related to overall job performance was unsupported. The hypothesis was tested using the facets of reasoning ability as well as a composite variable reflecting overall reasoning ability, and calculating the bivariate correlations of these reasoning variables with overall job performance ratings. Both Pearson’s and Spearman’s coefficients were calculated, and neither method yielded statistically significant results. This finding does not align with a significant body of previous research that has demonstrated that general mental ability is a predictor of job performance across job settings and roles (cf. McHenry et al., 1990; Ree et al., 1994; Schmidt & Hunter, 1998).

The research also explored the relationships between the reasoning ability variables and the facets of performance, using the factor score variables derived from the factor analysis of the performance data, representing the affective, cognitive, and citizenship domains of performance. All but one relationship was non-significant. A weak correlation ($\rho = .23$) significant at the $p < .05$ level was found between numerical reasoning and the first performance factor (affective). It is likely that this finding is non-meaningful due to the weak relationship and the fact that a clear relationship was not visible on the scatterplot. Additionally, there is no theoretical basis for a relationship between numerical reasoning and the affective domain of performance.

There are several possible explanations for the lack of relationships found between cognitive abilities and performance. One potential explanation is range restriction in the performance data. Although the overall performance ratings were approximately normally distributed, there was a complete absence of participants receiving an overall performance rating of 1 (‘unsatisfactory’). This can be taken as evidence of a slight range restriction. A restriction in range means that the full range of the criterion variable did not exist within the sample, and any potential relationship between ability and job performance may have been attenuated.

As the participants in the study were current employees, it is likely that performance would have been in the better range due to workplace performance management systems. Range restriction in the performance data may have also resulted from the way that participants were selected, for example, there might have been a bias towards nominating higher performing officers to participate in the study. The slight negative skew of ratings on the individual performance items (indicating ratings that were more favourable) is consistent with this explanation.

Further aspects of the study design and methodology may have contributed to the lack of relationship between ability and performance. One is the small number of participants. Further research could aim to explore the relationships between reasoning ability and performance using a larger sample size. Another possibility is that the current form of the performance appraisal tool may imply measurement properties that obscure the true nature of the relationship. A full validation of such a performance appraisal instrument is yet to be conducted for the local context. It cannot be excluded that a lack of relationship between ability and performance in this local study is partly due to artefacts relating to the measurement properties of data collection tools.

Criterion-related validity

The relationships between cognitive ability and job performance were analysed above in the context of hypothesis testing, as a substantive research issue. However, given the basic definition of criterion-related validity (Cohen et al., 2013; Murphy & Davidshofer, 2005), and the practical possibility that such ability tests may be used for selection purposes in the future, there is another context for interpretation.

This has to do with the properties of the GRT2 as a potential selection test for the given type of job. Once local job performance is accepted as a suitable choice for a criterion variable, the issue of whether GRT2 ability scores are related versus unrelated to this variable will serve as validation evidence. The correlation coefficients can be re-interpreted as evidence of criterion-related validity, and can be re-phrased as criterion-related validity coefficients.

Through testing the hypothesised relationship between cognitive ability and job performance, the study supplied some limited evidence toward levels of criterion-related validity for the GRT2 test in the corrections officer sample. The absence of a positive relationship between cognitive ability and performance implies that criterion-related validity for the GRT2 was unsupported in this study. Cognitive ability may be related to job performance in a corrections environment, but the GRT2 scores, in view of the correlation matrix produced, did not indicate an acceptable level of criterion-related validity of the GRT2 for performance.

A possible explanation for lack of criterion-related validity is the uniqueness of the corrections working environment and the nature of job itself. It is likely that officer performance requires person characteristics beyond reasoning abilities. If correlations with reasoning ability are as low as demonstrated here, the larger percentage of total variance in performance must be accounted for by other variables (i.e. variables outside reasoning ability, or at least outside the ability picked up by GRT2 scores). Interestingly, the factor analysis of performance data indicated that variance of performance in the corrections officer sample was dominated, to a major extent, by emotionally-oriented behaviours and attributes. It is theoretically plausible that cognitive ability would not predict these types of behaviours. Another form of ability, such as emotional intelligence, may be a better predictor. The finding that performance in the corrections officer sample is largely composed of emotional and social components is not surprising, given that an important facet of the role is working with people and the unique challenges of working with offenders. It is also plausible to assume that constructs entirely outside the realm of abilities, skills, and aptitudes (such as occupation-relevant personality traits) would predict a larger segment of variance in job performance. If so, the use of predictor tests outside ability tests will be promising.

(C) Participant characteristics, ability and performance

The only participant characteristic possible to explore in the data set was gender. *Hypothesis 4* was that there would be no gender difference in reasoning ability scores. The study revealed statistically significant differences between scores on two out of three reasoning dimensions. The hypothesis was supported for abstract reasoning, but not for verbal and numerical reasoning. Female participants scored significantly higher on the verbal reasoning subscale, whereas male participants scored significantly higher on the numerical reasoning subscale.

As abstract reasoning is conceptually the closest to general mental ability out of the three reasoning subtypes, the finding that there was no significant gender difference on this subscale is consistent with previous research that has suggested there are no gender differences in general mental ability. Previous research on the topic has revealed mixed findings in relation to more specific cognitive abilities. The present study is consistent with research that has found that males outperform females on assessments of numerical ability, and females do better on verbal ability, but it is inconsistent with reports that have revealed no significant differences in mathematical and verbal ability (cf. meta-analyses by Hyde et al., 1990; Hyde & Linn, 1988).

The relationships between the GRT2 subscales were explored separately for male and female participants. Some differences were found in the relationships between the ability dimensions when looking at the data for males and females separately. There was a stronger relationship between verbal reasoning and each of abstract and numerical reasoning for males. For females, there was a stronger relationship between abstract and numerical reasoning, in comparison to males. These findings indicate that the relationship between general and specific abilities may vary between genders, as well as the relationship between fluid and crystallised intelligences.

The study also examined the performance ratings for male and female participants. No significant differences were found, suggesting that male and female officers received similar overall ratings. However, there was an interesting trend when looking at ratings on individual items. Out of the 23 items in the performance measure, female participants were rated higher than male participants on 21 items. The differences were not statistically significant, however.

The final analysis looked at the relationship between cognitive ability and job performance (overall rating) separately for male and female participants. Close relationships were not found in these narrower groups either, consistent with the finding when exploring the data for the full group (male and female participants combined).

7.3. Limitations

For empirical research in organisational psychology, including that investigating relationships between ability and performance, a fundamental requirement is appropriate sampling. Best practice requires genuine probabilistic sampling, at least within a country and a pre-defined occupational or organisational environment. It is an essential requirement that the research tightly defines the relevant population and works with a sample as representative as possible of the population (Frankfort-Nachmias & Nachmias, 1996).

The current study, based on data sharing under special permission, used an existing data set provided. As a consequence, it was not possible for the researcher to follow best-practice procedures and define a sampling strategy before selection of a sample. Fortunately, the details of the original data collection were known to the researcher, and there is some evidence to support representativeness of the participant group.

It was possible to clearly define a finite population for this study (i.e. corrections officers working in New Zealand at a given time). However, it was not possible to assess the full population during the original data collection, and a sample was selected (i.e. the 88 officers in the study). Staff from seven of the seventeen prison sites were represented in the sample. In the original data collection, attempts were made to ensure that the sample represented the population as much as possible. For example, managers were asked to nominate corrections officers representing different levels of performance, and the ethnic and gender composition of the corrections officer workforce.

Although the sample in the present study can be considered representative of the population, it cannot, conservatively speaking, be considered a ‘true sample’. A true sample would have been obtained through a systematic sampling strategy (Sampath, 2005). A sampling strategy or design defines the strategy for selecting a subset from the population, for example, through stratified sampling and probability sampling (Frankfort-Nachmias & Nachmias, 1996; Hansen, Hurwitz, & Madow, 1993; Sampath, 2005). If simple probability sampling had been used, then each possible subset of the population would have had equal chance of being selected as the sample (Sampath, 2005). A sampling strategy would have provided greater assurance that the sample matched the population as much as possible in terms of person characteristics.

The use of non-probability sampling techniques, such as convenience samples and purposive samples (Frankfort-Nachmias & Nachmias, 1996), is common in organisational psychology. There are a number of reasons why non-probability sampling was used during the data collection, for example, cost, feasibility, and minimising disruption to staff and workplace routines.

The limitations of the sampling strategy must be taken into account when considering the generalisability of these results to the wider corrections environment and to wider environments and occupational settings.

7.4. Recommendations and directions for future research

While overcoming these limitations within the confines of the project was not logistically feasible, the author notes the avenues of improving this type of organisational research and conducting extended and improved versions of the project. Several recommendations are made for future research. The factor analysis revealed that the affective or emotional domain, along with self-awareness, explains a large proportion of variance in performance in the corrections context. This has implications for selection procedures. It is recommended to further explore psychometric methods to support such selection decisions (for example, through broadening or changing the predictor constructs, and opting for different predictor tests, such as those measuring emotional intelligence or occupational personality traits).

Another area for investigation stems from the acknowledgement that a corrections context is unique. Some organisations have had success using occupational specific or contextualised assessments. For Corrections, this would mean contextualising the items on the assessment. This of course would take significant investment, whereas the benefits of using a commercially available test are also clear.

Another recommendation is further refinement of the performance appraisal tool. The factor analysis of the performance data was strictly an exploratory, initial analysis. It revealed one way of conceptualising performance in this context. A new performance measure could be more consciously built around these identified dimensions (affective, cognitive, citizenship), with separate subscales for each dimension of performance. The tool would require psychometric development, including re-estimation of reliability, validation in the Corrections context, especially for construct validity, and more advanced versions of item analysis.

Although criterion-related validation of the cognitive ability test was not achieved in this setting, there are several pathways to explore validation further, and produce more promising criterion-related validity coefficients. Most reasons for lack of validity have been mentioned already, for example, the small *N*, range restriction, and unreliability or other gaps regarding the criterion measurement (assessment of job performance). However, it might also be the case that there is something unique or different about a corrections context that means that personal characteristics beyond cognitive ability are predictors of performance in this setting.

It is important to note that he cognitive ability assessment is just one stage of a regular, multi-stage selection procedure. It is to be used in conjunction with other methods (e.g. assessment centre, interview, and other psychometric tests).

While the tool was not validated as such, information was collected about its reliability that adds to the evidence base about the psychometric properties of the GRT2. It also appears to be suitable for use in this population (i.e. pitched at the right level).

The results have some generalisability. Although the group sizes for men and women were unequal, the gender composition of the combined sample reflected the population of corrections officers, which is a predominantly male occupation. Therefore, the findings from the analyses conducted with the full sample have increased generalisability to the wider corrections officer population. The findings could also be of relevance to other occupations that are similar, for example, the police where use of psychometric tests in officer selection is a well-known option.

APPENDICES

Appendix A

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227-257. doi: 10.1016/S0160-2896(96)90016-1

Ackerman, P. L., & Beier, M. E. (2003). Trait complexes, cognitive investment, and domain knowledge. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 1-30). Cambridge, England: Cambridge University Press.

Anderson, N., & Herriot, P. (Eds.). (1997). *International handbook of selection and assessment*. Chichester, England: John Wiley & Sons.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52. doi: 10.1177/0093854890017001004

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28*, 369-404. doi: 10.1111/j.1745-9125.1990.tb01330.x

Antonio, M. E., Young, J. L., & Winegard, L. M. (2009). Reinforcing positive behavior in a prison: Whose responsibility is it? *Journal of Offender Rehabilitation, 48*, 53-66. doi: 10.1080/10509670802572284

Antonowicz, D. H., & Ross, R. R. (1994). Essential components of successful rehabilitation programs for offenders. *International Journal of Offender Therapy and Comparative Criminology, 38*, 97-104. doi: 10.1177/0306624X9403800202

Atkinson, D. F., & Mann, R. E. (2012). Prison officers as observers of offence paralleling behaviours: An untapped resource? *Journal of Forensic Psychiatry & Psychology, 23*, 139-155. doi: 10.1080/14789949.2012.668209

Bateman, T. S., & Organ, D. W. (1983). Job satisfaction and the good soldier: The relationship between affect and employee "citizenship". *Academy of Management Journal, 26*, 587-595. doi: 10.2307/255908

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences, 30*, 977-994. doi: 10.1016/S0191-8869(00)00087-8

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences, 11*, 169-232. doi: 10.1017/S0140525X00049244

Blackwell, L. S., Rodriguez, S., & Guerra-Carrillo, B. (2015). Intelligence as a malleable construct. In S. Goldstein, D. Princiotta & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 263-282). New York, NY: Springer.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*, 383-405. doi: 10.1076/jcen.24.3.383.981

Bonta, J., Bourgon, G., Rugge, T., Scott, T., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior, 38*, 1127-1148. doi: 10.1177/0093854811420678

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisory experience in supervisor performance. *Journal of Applied Psychology, 78*, 443-449. doi: 10.1037/0021-9010.78.3.443

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*, 99-109. doi: 10.1207/s15327043hup1002_3

Bourgon, G., Gutierrez, L., & Ashton, J. (2011). The evolution of community supervision practice: The transformation from case manager to change agent. *Irish Probation Journal, 8*, 28-48.

Campbell , D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. doi: 10.1037/h0046016

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol 1* (2nd ed., pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Carson, J. (2014). Mental testing in the early twentieth century. *History of Psychology, 17*, 249-255. doi: 10.1037/a0037475

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22. doi: 10.1037/h0046743

Cattell, R. B. (Ed.). (1987). *Intelligence: Its structure, growth and action*. Amsterdam, Netherlands: Elsevier Science.

Chamorro-Premuzic, T., & Furnham, A. (2004). A possible model for understanding the personality-intelligence interface. *British Journal of Psychology, 95*, 249-264. doi: 10.1348/000712604773952458

Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and intellectual competence*. Mahwah, NJ: Lawrence Erlbaum.

Chamorro-Premuzic, T., & Furnham, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology, 10*, 251-267. doi: 10.1037/1089-2680.10.3.251

Chen, M. J. (1994). Chinese and Australian concepts of intelligence. *Psychology and Developing Societies, 6*, 103-117. doi: 10.1177/097133369400600202

Cleveland, J. N., & Colella, A. (2010). Criterion validity and criterion deficiency: What we measure well and what we ignore. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 551-567). New York, NY: Routledge.

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: An introduction to tests and measurement* (8th ed.). New York, NY: McGraw-Hill.

Colom, R., Juan-Espinosa, M., Abad, F., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence, 28*, 57-68. doi: 10.1016/s0160-2896(99)00035-5

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104. doi: 10.1037/0021-9010.78.1.98

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: HarperCollins.

Cronbach, L. J., & Drenth, P. J. D. (1973). *Mental tests and cultural adaptation*. The Hague, Netherlands: Mouton.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. doi: 10.1037/h0040957

Cuéllar, I., & Paniagua, F. A. (2000). *Handbook of multicultural mental health: Assessment and treatment of diverse populations*. San Diego, CA: Academic Press.

Cullen, F. T., & Gendreau, P. (2001). From nothing works to what works: Changing professional ideology in the 21st century. *Prison Journal, 81*, 313-338. doi: 10.1177/0032885501081003002

Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology*. Boston, MA: Allyn and Bacon.

Dana, R. H. (2000). *Handbook of cross-cultural and multicultural personality assessment. Personality and clinical psychology series.* Mahwah, NJ: Lawrence Erlbaum.

Department of Corrections. (2007). *The effectiveness of correctional treatment*. Wellington, New Zealand: Author. Retrieved from http://www.corrections.govt.nz/resources/the-effectiveness-of-correctional-treatment.html

Department of Corrections. (2009). *What works now? A review and update of research evidence relevant to offender rehabilitation practices within the Department of Corrections*. Wellington, New Zealand: Author.

Department of Corrections. (2012). *Corrections officer job description*. Wellington, New Zealand: Author.

Department of Corrections. (2013). *Trends in the offender population*. Wellington, New Zealand: Author.

Department of Corrections. (2014). *Statement of intent: 2014 to 2018*. Wellington, New Zealand: Author.

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol 1* (2nd ed., pp. 577-636). Palo Alto, CA: Consulting Psychologists Press.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi: 10.1037/1082-989x.4.3.272

Farkas, M. A. (1999). Correctional officer attitudes toward inmates and working with inmates in a "get tough" era. *Journal of Criminal Justice, 27*, 495-506. doi: 10.1016/S0047-2352(99)00020-3

Farkas, M. A., & Manning, P. K. (1997). The occupational culture of corrections and police officers. *Journal of Crime and Justice, 20*, 51-68. doi: 10.1080/0735648X.1997.9721581

Finney, C., Stergiopoulos, E., Hensel, J., Bonato, S., & Dewa, C. S. (2013). Organizational stressors associated with job stress and burnout in correctional officers: A systematic review. *BMC Public Health, 13*(82), 1-13. doi: 10.1186/1471-2458-13-82

Fleishman, E. A. (1967). Development of a behavior taxonomy for describing human tasks: A correlational-experimental approach. *Journal of Applied Psychology, 51*, 1-10. doi: 10.1037/h0024073

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist, 27*, 1017-1032. doi: 10.1037/h0033881

Fleishman, E. A. (1975). Toward a taxonomy of human performance. *American Psychologist, 30*, 1127-1149. doi: 10.1037/0003-066x.30.12.1127

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences* (5th ed.). New York, NY: St Martins Press.

Furnham, A. (2008). Intelligence and cognitive abilities at work. In S. Cartwright & G. L. Cooper (Eds.), *The Oxford handbook of personnel psychology* (pp. 7-36). New York, NY: Oxford University Press.

Furnham, A., Dissou, G., Sloan, P., & Chamorro-Premuzic, T. (2007). Personality and intelligence in business people: A study of two personality and two intelligence measures. *Journal of Business and Psychology, 22*, 99-109. doi: 10.1007/s10869-007-9051-z

Gardner, H. (1998). A multiplicity of intelligences. *Scientific American Presents: Exploring Intelligence, 9*(4), 19-23.

Gardner, H. (2006). *Multiple intelligences: New horizons*. New York, NY: Basic Books.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher, 18*, 4-10. doi: 10.2307/1176460

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2015). *Human resource selection* (8th ed.). Boston, MA: Cengage.

Gendreau, P. (1996). Offender rehabilitation: What we know and what needs to be done. *Criminal Justice and Behavior, 23*, 144-161. doi: 10.1177/0093854896023001010

Gendreau, P., & Keyes, D. (2001). Making prisons safer and more humane environments. *Canadian Journal of Criminology, 43*, 123-130.

Gendreau, P., & Ross, B. (1979). Effective correctional treatment: Bibliotherapy for cynics. *Crime & Delinquency, 25*, 463-489. doi: 10.1177/001112877902500405

Gilbert, M. J. (1997). The illusion of structure: A critique of the classical model of organization and the discretionary power of correctional officers. *Criminal Justice Review, 22*, 49-64. doi: 10.1177/073401689702200105

Goldstein, S., Princiotta, D., & Naglieri, J. A. (Eds.). (2015). *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*. New York, NY: Springer.

Goleman, D. (1995). *Emotional intelligence*. New York, NY: Bantam Books.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*, 79-132. doi: 10.1016/S0160-2896(97)90014-3

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents, 9*(4), 24-29.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*, 1115-1124. doi: 10.1037/0003-066x.52.10.1115

Gregory, R. J. (2011). *Psychological testing: History, principles, and applications* (6th ed.). Boston, MA: Allyn & Bacon.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: John Wiley & Sons.

Gruszka, A., Matthews, G., & Szymura, B. (Eds.). (2010). *Handbook of individual differences in cognition: Attention, memory, and executive control*. New York, NY: Springer.

Guenole, N., Englert, P., & Taylor, P. J. (2003). Ethnic group differences in cognitive ability test scores within a New Zealand applicant sample. *New Zealand Journal of Psychology, 32*(1), 49-54.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179-203. doi: 10.1016/0160-2896(84)90008-4

Hambleton, R. K. (1996). *Guidelines for adapting educational and psychological tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164-172. doi: 10.1027//1015-5759.17.3.164

Hambleton, R. K. (Ed.). (1983). *Applications of item response theory*. Vancouver, Canada: Educational Research Institute of British Columbia.

Hambleton, R. K., & De Jong, J. H. A. L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing, 20*, 127-134. doi: 10.1191/0265532203lt247xx

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1993). *Sample survey methods and theory* (Vol. 1). New York, NY: Wiley.

Hattrup, K., Schmitt, N., & Landis, R. S. (1992). Equivalence of constructs measured by job-specific and commercially available aptitude tests. *Journal of Applied Psychology, 77*, 298-308. doi: 10.1037/0021-9010.77.3.298

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*, 191-205. doi: 10.1177/1094428104263675

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45. doi: 10.1126/science.7604277

Hepburn, J. R., & Knepper, P. E. (1993). Correctional officers as human services workers: The effect on job satisfaction. *Justice Quarterly, 10*, 315-337. doi: 10.1080/07418829300091841

Hesketh, B., & Griffin, B. (2008). Selection and training for work adjustment and adaptability. In S. Cartwright & G. L. Cooper (Eds.), *The Oxford handbook of personnel psychology* (pp. 366-387). New York, NY: Oxford University Press.

Hobbs, G. S., & Dear, G. E. (2000). Prisoners' perceptions of prison officers as sources of support. *Journal of Offender Rehabilitation, 31*, 127-142. doi: 10.1300/J076v31n01_09

Horn, J. L. (1965a). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois.

Horn, J. L. (1965b). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185. doi: 10.1007/BF02289447

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*, 242-259. doi: 10.1037/h0025662

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73-98). New York, NY: Guilford Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253-270. doi: 10.1037/h0023816

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129. doi: 10.1016/0001-6918(67)90011-X

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology, 94*, 1128-1145. doi: 10.1037/a0015978

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362. doi: 10.1016/0001-8791(86)90013-8

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98. doi: 10.1037/0033-2909.96.1.72

Hutchinson, T. P. (1991). *Controversies in item response theory*. Adelaide, Australia: Rumsby Scientific Publishing.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155. doi: 10.1037/0033-2909.107.2.139

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69. doi: 10.1037/0033-2909.104.1.53

Jensen, A. R. (1986). *g*: Artifact or reality? *Journal of Vocational Behavior, 29*, 301-331. doi: 10.1016/0001-8791(86)90011-4

Jensen, A. R. (1987). Psychometric *g* as a focus of concerted research effort. *Intelligence, 11*, 193-198. doi: 10.1016/0160-2896(87)90005-5

Jensen, A. R. (1992). Understanding *g* in terms of information processing. *Educational Psychology Review, 4*, 271-308. doi: 10.1007/bf01417874

Jensen, A. R. (2002). Galton's legacy to research on intelligence. *Journal of Biosocial Science, 34*, 145-172. doi: 10.1017/s0021932002001451

Johnson, W., & Bouchard, T. J. (2007). Sex differences in mental abilities: *g* masks the dimensions on which they lie. *Intelligence, 35*, 23-39. doi: 10.1016/j.intell.2006.03.012

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*, 376-407. doi: 10.1037/0033-2909.127.3.376

Kline, P. (1994). *An easy guide to factor analysis*. London, England: Routledge.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London, England: Routledge.

Kline , T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage Publications.

Kranzler, J. H., Keith, T. Z., & Flanagan, D. P. (2000). Independent examination of the factor structure of the Cognitive Assessment System (CAS): Further evidence challenging the construct validity of the CAS. *Journal of Psychoeducational Assessment, 18*, 143-159. doi: 10.1177/073428290001800204

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148-161. doi: 10.1037/0022-3514.86.1.148

Lambert, E. G., Hogan, N. L., & Barton, S. M. (2001). The impact of job satisfaction on turnover intent: A test of a structural measurement model using a national sample of workers. *Social Science Journal, 38*, 233-250. doi: 10.1016/S0362-3319(01)00110-0

Lambert, E. G., & Paoline, E. A. (2008). The influence of individual, job, and organizational characteristics on correctional staff job stress, job satisfaction, and organizational commitment. *Criminal Justice Review, 33*, 541-564. doi: 10.1177/0734016808320694

Lang, J. W. B., Kersting, M., Hülsheger, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology, 63*, 595-640. doi: 10.1111/j.1744-6570.2010.01182.x

Lievens, F., & Reeve, C. L. (2012). Where I-O psychology should really (re)start its investigation of intelligence constructs and their measurement. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 5*, 153-158. doi: 10.1111/j.1754-9434.2012.01421.x

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498. doi: 10.2307/1130467

Lipton, D., Martinson, R., & Wilks, J. (1975). *The effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York, NY: Praeger.

Lombardo, L. X. (1989). *Guards imprisoned: Correctional officers at work* (2nd ed.). Cincinnati, OH: Anderson Publishing Co.

Lord, F. (1959). An approach to mental test theory. *Psychometrika, 24*, 283-302. doi: 10.1007/BF02289812

Lösel, F. (1995). Increasing consensus in the evaluation of offender rehabilitation? Lessons from recent research syntheses. *Psychology, Crime & Law, 2*, 19-39. doi: 10.1080/10683169508409762

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) '''general intelligence,' objectively determined and measured''. *Journal of Personality and Social Psychology, 86*, 96-111. doi: 10.1037/0022-3514.86.1.96

Lyman, H. B. (1998). *Test scores and what they mean* (6th ed.). Boston, MA: Allyn and Bacon.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95-107. doi: 10.1037/h0056029

Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest, 35*, 22-54.

Mayer, J. D., Caruso, D. R., Panter, A. T., & Salovey, P. (2012). The growing significance of hot intelligences. *American Psychologist, 67*, 502-503. doi: 10.1037/a0029456

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298. doi: 10.1016/s0160-2896(99)00016-1

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17*, 433-442. doi: 10.1016/0160-2896(93)90010-3

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*, 197-215.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1*, 232-242. doi: 10.1037/1528-3542.1.3.232

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

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. doi: 10.1111/j.1744-6570.1990.tb01562.x

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*, 71-83. doi: 10.1207/s15327043hup1002_1

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480. doi: 10.1037/0021-9010.79.4.475

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Mulvey, E. P., Arthur, M. W., & Reppucci, N. D. (1993). The prevention and treatment of juvenile delinquency: A review of the research. *Clinical Psychology Review, 13*, 133-167. doi: 10.1016/0272-7358(93)90038-N

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than *g*. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3-30). San Francisco, CA: Jossey-Bass Publishers.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Naglieri, J. A., & Das, J. P. (1990). Planning, attention, simultaneous, and successive (PASS) cognitive processes as a model for intelligence. *Journal of Psychoeducational Assessment, 8*, 303-337. doi: 10.1177/073428299000800308

Naglieri, J. A., Das, J. P., & Goldstein, S. (2012). Planning, attention, simultaneous, successive: A cognitive-processing-based theory of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 178-194). New York, NY: Guilford Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101. doi: 10.1037/0003-066x.51.2.77

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*, 130-159. doi: 10.1037/a0026699

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*, 1-18. doi: 10.1016/0022-2496(66)90002-2

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

O'Fallon, C. (2014). Translating the right relationship into right practice: Right Track in NZ prisons. *Practice – The New Zealand Corrections Journal, 2*(1), 42-44.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*, 845-851. doi: 10.1037/0021-9010.79.6.845

Ones, D., Dilchert, S., & Viswesvaran, C. (2012). Cognitive abilities. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 179-224). New York, NY: Oxford University Press.

Ones, D., Dilchert, S., Viswesvaran, C., & Salgado, J. (2010). Cognitive abilities. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 255-275). New York, NY: Routledge.

OPRA Consulting Group. (2009). Practice ability assessment. Retrieved from http://www.opragroup.com/images/opra/flash/OPRAAbility.swf

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10*, 85-97. doi: 10.1207/s15327043hup1002_2

Organ, D. W., Podsakoff, N. P., & MacKenzie, S. B. (2006). *Organizational citizenship behavior: Its nature, antecedents, and consequences*. Thousand Oaks, CA: Sage.

Palmer, T. (1975). Martinson revisited. *Journal of Research in Crime and Delinquency, 12*, 133-152. doi: 10.1177/002242787501200206

Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science 19*, 957-961. doi: 10.1111/j.1467-9280.2008.02182.x

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15*, 425-448. doi: 10.1002/per.416

Philliber, S. (1987). Thy brother's keeper: A review of the literature on correctional officers. *Justice Quarterly, 4*, 9-37. doi: 10.1080/07418828700089171

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology, 51*, 390-395. doi: 10.1037/0022-006x.51.3.390

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102-1114. doi: 10.1037/0003-066X.47.9.1102

Psytech International Ltd. (n.d). General and graduate reasoning tests. Bedfordshire, England: Psytech International Ltd.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*, 612-624. doi: 10.1037//0021-9010.85.4.612

Pulakos, E. D., Mueller-Hanson, R., & Nelson, J. (2012). Adaptive performance and trainability as criteria in selection research. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 595-613). New York, NY: Oxford University Press.

Ree, M. J., & Earles, J. A. (1991). The stability of *g* across different methods of estimation. *Intelligence, 15*, 271-278. doi: 10.1016/0160-2896(91)90036-d

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79*, 518-524. doi: 10.1037/0021-9010.79.4.518

Reeve, C. L., & Hakel, M. D. (2002). Asking the right questions about *g*. *Human Performance, 15*, 47-74. doi: 10.1080/08959285.2002.9668083

Reynolds, C. R., & Livingston, R. B. (2012). *Mastering modern psychological testing: Theory and methods* (1st ed.). Boston, MA: Pearson Education.

Robinson, D., Porporino, F. J., & Simourd, L. (1997). The influence of educational attainment on the attitudes and job performance of correctional officers. *Crime & Delinquency, 43*, 60-77. doi: 10.1177/0011128797043001004

Rodriguez, C. M., Treacy, L. A., Sowerby, P. J., & Murphy, L. E. (1998). Applicability of Australian adaptations of intelligence tests in New Zealand with a Dunedin sample of children. *New Zealand Journal of Psychology, 27*(1), 4-12.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group difference in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297-330. doi: 10.1111/j.1744-6570.2001.tb00094.x

Sackett, P. R. (2002). The structure of counterproductive work behaviors: Dimensionality and relationships with facets of job performance. *International Journal of Selection and Assessment, 10*, 5-11. doi: 10.1111/1468-2389.00189

Sackett, P. R. (2007). Revisiting the origins of the typical-maximum performance distinction. *Human Performance, 20*, 179-185. doi: 10.1080/08959280701332968

Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczo, R. M. (2006). Citizenship and counterproductive behavior: Clarifying relations between the two domains. *Human Performance, 19*, 441-464. doi: 10.1207/s15327043hup1904_7

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112-118. doi: 10.1037//0021-9010.85.1.112

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*, 482-486. doi: 10.1037/0021-9010.73.3.482

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56*, 573-605. doi: 10.1111/j.1744-6570.2003.tb00751.x

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9*, 185-211. doi: 10.2190/dugg-p24e-52wk-6cdg

Sampath, S. (2005). *Sampling theory and methods* (2nd ed.). Harrow, England: Alpha Science International Ltd.

Samuda, R. J. (Ed.). (1998). *Advances in cross-cultural assessment*. Thousand Oaks, CA: Sage Publications.

Samuda, R. J., & Wolfgang, A. (Eds.). (1985). *Intercultural counselling and assessment: Global perspectives*. Lewiston, NY: C.J. Hogrefe.

Scherbaum, C. A., Goldstein, H. W., Yusko, K. P., Ryan, R., & Hanges, P. J. (2012). Intelligence 2.0: Reestablishing a research program on *g* in I-O psychology. *Industrial and Organizational Psychology, 5*, 128-148. doi: 10.1111/j.1754-9434.2012.01419.x

Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science, 1*, 89-92. doi: 10.1111/1467-8721.ep10768758

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*, 262-274. doi: 10.1037/0033-2909.124.2.262

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

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353. doi: 10.1037/1040-3590.8.4.350

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 45-65. doi: 10.1146/annurev-orgpsych-031413-091255

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: Guilford Press.

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review, 25*, 12-27. doi: 10.1016/j.hrmr.2014.09.004

Schuerger, J. M., Kochevar, K. F., & Reinwald, J. E. (1982). Male and female corrections officers: Personality and rated performance. *Psychological Reports, 51*, 223-228. doi: 10.2466/pr0.1982.51.1.223

Schweizer, K., Troche, S. J., & Rammsayer, T. H. (2011). On the special relationship between fluid and general intelligence: New evidence obtained by considering the position effect. *Personality and Individual Differences, 50*, 1249-1254. doi: 10.1016/j.paid.2011.02.019

Shultz, K. S., & Whitney, D. J. (2005). *Measurement theory in action: Case studies and exercises*. Thousand Oaks, CA: Sage Publications.

Shum, D., O'Gorman, J., & Myors, B. (2006). *Psychological testing and assessment*. Melbourne, Australia: Oxford University Press.

Shusman, E. J., Inwald, R. E., & Landa, B. (1984). Correction officer job performance as predicted by the IPI and MMPI: A validation and cross-validation study. *Criminal Justice and Behavior, 11*, 309-329. doi: 10.1177/0093854884011003004

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*, 201-292. doi: 10.2307/1412107

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. London, England: Macmillan.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*, 950-958. doi: 10.1037/0003-066x.60.9.950

Sproule, C. F., & Berkley, S. (2001). The selection of entry-level corrections officers: Pennsylvania research. *Public Personnel Management, 30*, 377-418. doi: 10.1177/009102600103000307

Stein, S. J., & Deonarine, J. M. (2015). Current concepts in the assessment of emotional intelligence. In S. Goldstein, D. Princiotta & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 381-402). New York, NY: Springer.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Oxford, England: Lawrence Erlbaum.

Sternberg, R. J. (1983). Components of human intelligence. *Cognition, 15*, 1-48. doi: 10.1016/0010-0277(83)90032-X

Sternberg, R. J. (1984a). Toward a triarchic theory of human intelligence. *Behavioral and Brain Sciences, 7*, 269-315. doi: 10.1017/s0140525x00044629

Sternberg, R. J. (1984b). What should intelligence tests test? Implications of a triarchic theory of intelligence for intelligence testing. *Educational Researcher, 13*, 5-15. doi: 10.2307/1174622

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*, 325-338. doi: 10.1037/0003-066x.59.5.325

Sternberg, R. J. (2012). The triarchic theory of successful intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 156-177). New York, NY: Guilford Press.

Sternberg, R. J. (2014). Teaching about the nature of intelligence. *Intelligence, 42*, 176-179. doi: 10.1016/j.intell.2013.08.010

Sternberg, R. J. (Ed.). (1994). *Encyclopedia of human intelligence*. New York, NY: Macmillan.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge, England: Cambridge University Press.

Sternberg, R. J., & Kaufman, J. C. (1996). Innovation and intelligence testing: The curious case of the dog that didn't bark. *European Journal of Psychological Assessment, 12*, 175-182. doi: 10.1027/1015-5759.12.3.175

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49*, 479-502. doi: 10.1146/annurev.psych.49.1.479

Sternberg, R. J., & Kaufman, S. B. (2012). Trends in intelligence research. *Intelligence, 40*, 235-236. doi: 10.1016/j.intell.2012.01.007

Sternberg, R. J., & Kaufman, S. B. (Eds.). (2011). *The Cambridge handbook of intelligence*. Cambridge, England: Cambridge University Press.

Sternberg, R. J., Lautrey, J., & Lubart, T. I. (Eds.). (2003). *Models of intelligence: International perspectives*. Washington, DC: American Psychological Association.

Sternberg, R. J., & Pretz, J. E. (Eds.). (2005). *Cognition and intelligence: Identifying the mechanisms of the mind*. Cambridge, England: Cambridge University Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680. doi: 10.1126/science.103.2684.677

Suzuki, L. A., Meller, P. J., & Ponterotto, J. G. (Eds.). (1996). *Handbook of multicultural assessment: Clinical, psychological, and educational applications*. San Francisco, CA: Jossey-Bass.

Suzuki, L. A., & Ponterotto, J. G. (Eds.). (2008). *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (3rd ed.). San Francisco, CA: Jossey-Bass.

Terman, L. M. (1921). Intelligence and its measurement: A symposium--II. *Journal of Educational Psychology, 12*, 127-133. doi: 10.1037/h0064940

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (1st ed.). Washington, DC: American Psychological Association.

Thorndike, E. L. (1921). Intelligence and its measurement: A symposium--I. *Journal of Educational Psychology, 12*, 124-127. doi: 10.1037/h0064596

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

Udechukwu, I. I. (2009). Correctional officer turnover: Of Maslow's needs hierarchy and Herzberg's motivation theory. *Public Personnel Management, 38*, 69-82. doi: 10.1177/009102600903800205

Valentin Kvist, A., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36*, 422-436. doi: 10.1016/j.intell.2007.08.004

van de Vijver, F. (2002). Test adaptation/translation methods. In R. Fernández Ballesteros (Ed.), *Encyclopedia of psychological assessment* (Vol. 2, pp. 960-964). London, England: Sage Publications Ltd.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*, 842-861. doi: 10.1037/0033-295X.113.4.842

Van Hein, J. L., Kramer, J. J., & Hein, M. (2007). The validity of the Reid Report for selection of corrections staff. *Public Personnel Management, 36*, 269-280. doi: 10.1177/009102600703600306

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006a). Beyond *g*: Putting multiple intelligences theory to the test. *Intelligence, 34*, 487-502. doi: 10.1016/j.intell.2006.02.004

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006b). *g* and the measurement of multiple intelligences: A response to Gardner. *Intelligence, 34*, 507-510. doi: 10.1016/j.intell.2006.04.006

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*, 412-423. doi: 10.1016/j.intell.2010.04.006

Walkey, F. H., & Welch, G. (2010). *Demystifying factor analysis: How it works and how to use it*. Bloomington, IN: Xlibris Corporation.

Warm, T. A. (1978). *A primer of item response theory* (No. Technical Report 940279). Oklahoma City, OK: Coast Guard Institute.

Whitehead, P. (2014). Therapeutic communities: From a programme in prison to prison as the programme. *Practice – The New Zealand Corrections Journal, 2*(2), 47-53.

Wicks, R. J. (1974). *Applied psychology for law enforcement and correction officers*. New York, NY: McGraw-Hill

Wilhelm, O., & Engle, R. W. (Eds.). (2005). *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications.

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*, 601-617. doi: 10.1177/014920639101700305

Woehr, D. J., & Roch, S. (2012). Supervisory performance ratings. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 517-531). New York, NY: Oxford University Press.

Wolf, A., & Little, A. (Eds.). (1996). *Assessment in transition: Learning, monitoring and selection in an international perspective*. New York, NY: Pergamon.

Yang, S.-B., Brown, G. C., & Moon, B. (2011). Factors leading to corrections officers' job satisfaction. *Public Personnel Management, 40*, 359-369. doi: 10.1177/009102601104000407

Yang, S., & Sternberg, R. J. (1997). Conceptions of intelligence in ancient Chinese philosophy. *Journal of Theoretical and Philosophical Psychology, 17*, 101-119. doi: 10.1037/h0091164

Young, L. (2013). The right relationship. *Practice – The New Zealand Corrections Journal, 1*(1), 24-26.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442. doi: 10.1037/0033-2909.99.3.432

Appendix B

Information and consent form

CANDIDATE INFORMATION AND RELEASE FORM

Thank you for participating in this research. Your help today is critical to selecting staff that you would be happy to have standing beside you as you go about your day.

Why have I been asked to do this?

The purpose of today's session is to see how well three new assessments can identify the attributes needed to do well as a Corrections Officer. The three assessments are:

- A cognitive ability test that measure three separate cognitive abilities
- An emotional intelligence questionnaire
- A written work sample exercise

We are trialling a new cognitive ability test for use in selecting Corrections Officers (COs). An important part of introducing a test of this nature is to confirm its usefulness in identifying who does and who does not have the right attributes to be an effective CO. This involves carrying out some research on existing staff that will ultimately determine what scores on the tests applicants for CO roles will need to meet. Once we use the test in selection, candidate scores can be compared against the CO group aggregated scores (the norm group) so that we can decide whether the candidates have what it takes to be a CO.

After speaking with frontline prison staff about the key requirements to be an effective CO, we decided it would be useful to assess at selection how well applicants manage difficult people, particularly where people become aggressive and confrontational. This is a key aspect of keeping safe. The Staff Safety Action Plan contains an action on ensuring our selection approach results in new employees who you feel can be relied on as safe team members. So, as part of updating our selection approach we are looking at introducing a short assessment that measures emotional intelligence, or how well someone manages themselves and others generally and in difficult situations.

Finally we also know from speaking with frontline prison staff that the ability to write clear file notes and reports is critical to managing offenders' needs well. So we are looking at introducing a written test into the selection process.

How long is this going to take?

In total, the three assessments should take no longer than 1.5 hours to complete.

What's going to happen to my results?

We have engaged an external provider, OPRA Consulting Group, to help us with analyzing the data from the research. National Office Recruitment staff have been enlisted to provide feedback to you. The only people that will see individual results are the external provider and National Office recruitment staff. Your scores on the assessments are compared against a large group of NZ respondents who have already taken the tests, so that comparisons can be made. From time to time the collective results will be used for approved research projects. In this instance, results are made anonymous and so will not personally identify any one.

The results from today will be cross-referenced with a questionnaire administered to your manager that asks how you go about the key requirements of the CO role. This enables us to look at the link between the three assessments and later performance in the role so we can determine the right scores to use at selection. No individual results will be made available to your managers.

Will I get to see my results?

You will have the opportunity to get feedback on your results if you would like it. This will be verbal feedback only as there are restrictions around giving written test information to people who are not trained in interpreting that information. Results are considered valid for around 12 months, after which time data is made anonymous.

If you would like feedback, this can be made available to you from **19 June 2013.** Please contact Senior HR Adviser, National Office on x 68320 or via email if you would like feedback.

Please read through the following information and sign if you agree to the conditions of the assessment:

The assessment testing process is to assist in determining the link between the assessments and how well staff carry out the CO role as well as establishing a group of aggregate results against which to compare candidate scores.

- I understand that my results will be used for approved research projects. Any information used for research purposes will be anonymous and will not personally identify me.

- I understand that OPRA Consulting Group (the test supplier) will receive the database containing my results, and that it will be stored in a collective database for statistical use only, with my name held strictly confidential. Published statistics will not be traceable back to me.

- In line with the Privacy Act of 1993, I understand that I am entitled to access my test results. However, I understand that due to ethical concerns regarding the release of un-interpreted psychometric data, I will be given a summary of my results in the form of verbal feedback rather than being given the results in detail. This will be offered from National Office recruitment staff.

- The information I provide will be secured against loss, unauthorised access, modification, disclosure, and misuse.
- I understand that emotional intelligence assessments include a measure of the extent to which the answers given are a true representation of myself.

☐ I have told the test administrator of any physical, health or other issue that may impact on my performance
☐ I have enough lighting in the testing room to complete the evaluation
☐ I have reading glasses/contact lenses if required
☐ I am aware of the approximate length of time this evaluation will take
☐ I am aware of the nature of the evaluations that I will be undertaking
☐ I am aware that I am not entitled to use a calculator or dictionary in the assessments unless instructed by the administrator
☐ I have switched off my mobile phone (if I have one with me)

If there is any reason why today is not an appropriate time for me to be undertaking the assessment (e.g. due to personal illness or stress), then I have let the test administrator know (or will do so as soon as possible).

I have read, understand, and agree with the information and requirements relating to my assessment session.

Name: ______________________________ **Date:** ____________________

Signature: ____

Contact: ____

Appendix C

Job appraisal form

Manager Rating Form

Thank you for getting involved in the CO recruitment revamp project. As part of this project, we need to have an accurate rating of CO on-the-job performance. In the attached questionnaire we have identified a number of work-related areas that are important for performance, and which we need you to rate each nominated CO from your team against.

This is a multiple-choice questionnaire, and uses a 1 to 5 rating scale. When completing this questionnaire, please answer each question as honestly and accurately as you can. We expect to receive a number of ratings for each CO, as each person is expected to have areas for development as well as areas of strength.

This exercise should take you no longer than 5 minutes of your time per person. All ratings will remain confidential to OPRA Consulting, and will not be shared with anyone else inside the Department.

Questionnaire Items:

In terms of overall job performance, I rate this CO's performance as ….	Unsatisfactory	Needs improvement	Meets Standard	Slightly Above Standard	Well Above Standard

Please rate how frequently this individual …

Goes above and beyond the requirements of their role	Almost Never	Occasionally	Sometimes	Often	Almost Always
Has a strong desire to learn and invest in their personal development	Almost Never	Occasionally	Sometimes	Often	Almost Always
Voluntarily assists others, even when there is no personal benefit	Almost Never	Occasionally	Sometimes	Often	Almost Always

Please rate the individual's skill in each area below …

Communicates clearly orally and in writing	Almost Never	Occasionally	Sometimes	Often	Almost Always
Communicates clearly with numbers and data	Almost Never	Occasionally	Sometimes	Often	Almost Always
Builds positive relationships, regardless of nationality, age, and gender	Almost Never	Occasionally	Sometimes	Often	Almost Always
Picks up new learning quickly and without the need for repeated explanation	Almost Never	Occasionally	Sometimes	Often	Almost Always
Takes the time to understand the needs of Maori	Almost Never	Occasionally	Sometimes	Often	Almost Always
Proactively integrates their understanding of Maori needs when relating to Maori	Almost Never	Occasionally	Sometimes	Often	Almost Always

Can bounce back quickly from setbacks	Almost Never	Occasionally	Sometimes	Often	Almost Always
Thinks ahead, plans in advance	Almost Never	Occasionally	Sometimes	Often	Almost Always
Explores alternative options prior to taking action	Almost Never	Occasionally	Sometimes	Often	Almost Always
Is a trustworthy individual; is a person of their word	Almost Never	Occasionally	Sometimes	Often	Almost Always
Pays attention to the environment and situational cues	Almost Never	Occasionally	Sometimes	Often	Almost Always
Is in tune with his/her own feelings, moods, and emotions	Almost Never	Occasionally	Sometimes	Often	Almost Always
Expresses his/her feelings to the right degree, to the right people, and at the right time	Almost Never	Occasionally	Sometimes	Often	Almost Always
Can tune in to what makes people tick and connect with others	Almost Never	Occasionally	Sometimes	Often	Almost Always
Will consider others' feelings when decision making	Almost Never	Occasionally	Sometimes	Often	Almost Always
Can effectively manage their own emotions, both positive and negative	Almost Never	Occasionally	Sometimes	Often	Almost Always
Can positively lift others' mood, feelings, and emotions; can lift energy	Almost Never	Occasionally	Sometimes	Often	Almost Always
Can remain calm and focused during periods of high stress and pressure	Almost Never	Occasionally	Sometimes	Often	Almost Always
Is considered by colleagues to be a team player demonstrating loyalty to the team	Almost Never	Occasionally	Sometimes	Often	Almost Always

Appendix D

Figures and tables

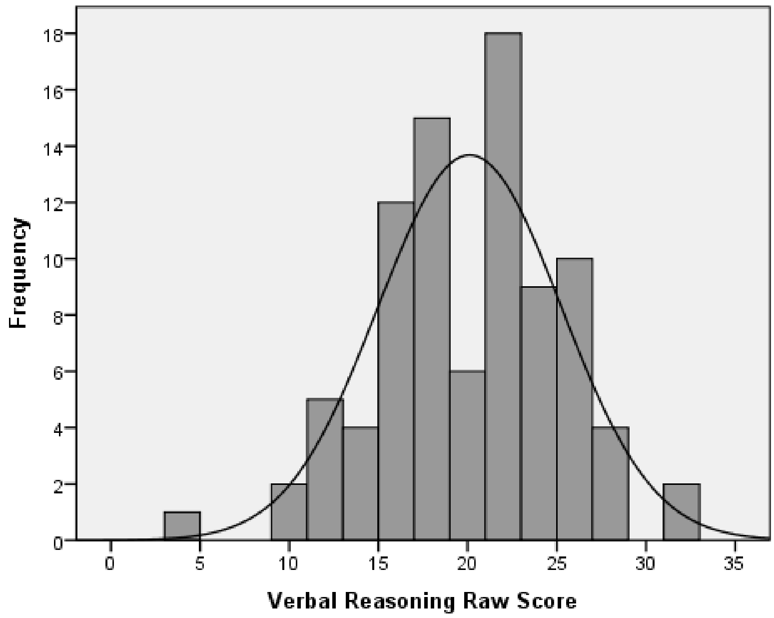


Figure A1. Distribution of GRT2 verbal reasoning scores
Note. *N* = 88, *M* = 19.67, *SD* = 5.10.

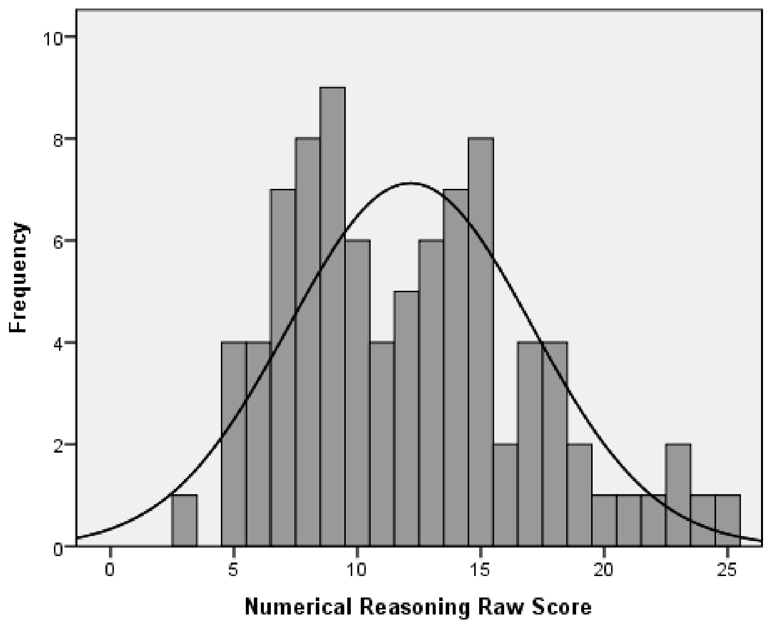


Figure A2. Distribution of GRT2 numerical reasoning scores
Note. *N* = 88, *M* = 12.15, *SD* = 4.93.

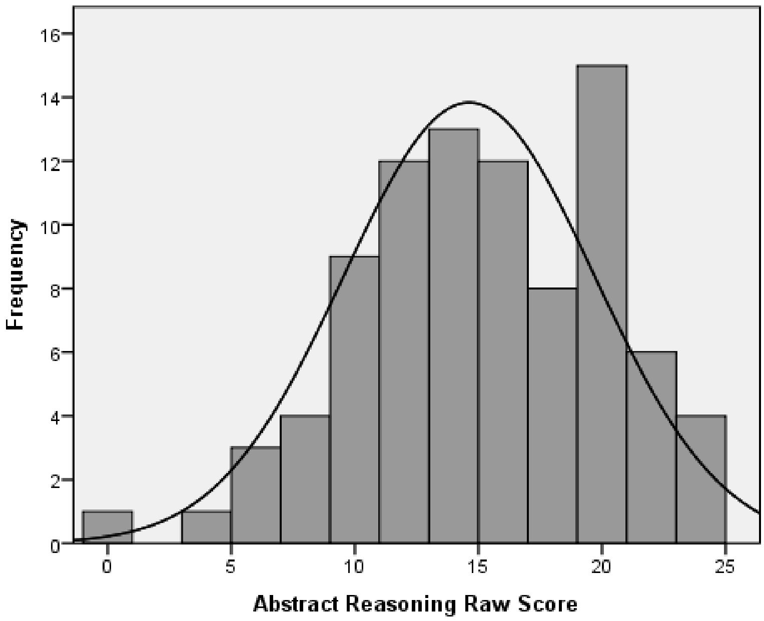


Figure A3. Distribution of GRT2 abstract reasoning scores
Note. *N* = 88, *M* = 14.62, *SD* = 5.08.

Table A1

Variable Names, Description, Possible Values, and Level of Measurement

Name	Concise description of variable	Possible values	Level of measurement
GEND	Participant gender	1 – 2 (M, F)	Nominal (dichotomous)
VR2	Verbal reasoning subscale raw score	0 – 35	Interval
NR2	Numerical reasoning subscale raw score	0 – 25	Interval
AR2	Abstract reasoning subscale raw score	0 – 25	Interval
OVERALL_GRT2	Overall GRT2 score (composite variable)	-3 – 3	Interval
PF1	Overall job performance [a]	1 – 5	Ordinal
PF2	Goes above and beyond the requirements of their role [b]	1 – 5	Ordinal
PF3	Has a strong desire to learn and invest in their personal development	1 – 5	Ordinal
PF4	Voluntarily assists others, even when there is no personal benefit	1 – 5	Ordinal
PF5	Communicates clearly orally and in writing	1 – 5	Ordinal
PF6	Communicates clearly with numbers and data	1 – 5	Ordinal
PF7	Builds positive relationships, regardless of nationality, age, and gender	1 – 5	Ordinal
PF8	Picks up new learning quickly and without the need for repeated explanation	1 – 5	Ordinal
PF9	Takes the time to understand the needs of Māori	1 – 5	Ordinal
PF10	Proactively integrates their understanding of Māori needs when relating to Māori	1 – 5	Ordinal
PF11	Can bounce back quickly from setbacks	1 – 5	Ordinal
PF12	Thinks ahead, plans in advance	1 – 5	Ordinal
PF13	Explores alternative options prior to taking action	1 – 5	Ordinal
PF14	Is a trustworthy individual; is a person of their word	1 – 5	Ordinal
PF15	Pays attention to the environment and situational cues	1 – 5	Ordinal
PF16	Is in tune with his/her own feelings, moods, and emotions	1 – 5	Ordinal
PF17	Expresses feelings to the right degree, to the right people, and at the right time	1 – 5	Ordinal

PF18	Can tune in to what makes people tick and connect with others	1 – 5	Ordinal
PF19	Will consider others' feelings when decision making	1 – 5	Ordinal
PF20	Can effectively manage their own emotions, both positive and negative	1 – 5	Ordinal
PF21	Can positively lift others' mood, feelings, and emotions; can lift energy	1 – 5	Ordinal
PF22	Can remain calm and focused during periods of high stress and pressure	1 – 5	Ordinal
PF23	Is considered by colleagues to be a team player demonstrating loyalty to the team	1 – 5	Ordinal
AFFECTIVE	Performance factor one (factor score composite variable)	-3 – 3	Interval
COGNITIVE	Performance factor two (factor score composite variable)	-3 – 3	Interval
CITIZENSHIP	Performance factor three (factor score composite variable)	-3 – 3	Interval
VerbalQ01 – VerbalQ35	Item responses for verbal reasoning subscale	0 – 1 [c] (Incorrect, Correct)	Nominal
NumericQ01 – NumericQ25	Item responses for numerical reasoning subscale	0 – 1 (Incorrect, Correct)	Nominal
AbstractQ01 – AbstractQ25	Item responses for abstract reasoning subscale	0 – 1 (Incorrect, Correct)	Nominal (dichotomous)

Note.
[a] PF1 (overall job performance) was a single item on the performance measure and not a sum or composite of other performance items. The item was rated as *1 = Unsatisfactory, 2 = Needs Improvement, 3 = Meets Standards, 4 = Slightly Above Standard, and 5 = Well Above Standard*

[b] PF 2 to 23 were rated as *1 = Almost Never, 2 = Occasionally, 3 = Sometimes, 4 = Often, 5 = Almost Always*

[c] The item-level data were received in a dichotomously coded format. The 'incorrect' category did not distinguish between items answered incorrectly and those that were not answered or 'timed out'.

Table A2

Mean, Median, Range, and Distribution Statistics for Performance Measure Items

Performance measure item		Mean	Median	*SD*	Min.	Max.	Skewness	*SE* of skewness	Kurtosis	*SE* of kurtosis
PF1	Overall job performance	3.33	3.00	.95	2	5	.12	.27	-.90	.54
PF2	Goes above and beyond the requirements of their role	3.19	3.00	1.06	1	5	-.13	.27	-.62	.54
PF3	Has a strong desire to learn and invest in their personal development	3.44	4.00	1.09	1	5	-.27	.27	-.88	.54
PF4	Voluntarily assists others, even when there is no personal benefit	3.50	4.00	1.17	1	5	-.42	.27	-.58	.54
PF5	Communicates clearly orally and in writing	3.68	4.00	.92	2	5	-.34	.27	-.63	.54
PF6	Communicates clearly with numbers and data	3.51	4.00	1.03	1	5	-.29	.27	-.80	.54
PF7	Builds positive relationships, regardless of nationality, age, and gender	3.71	4.00	.97	1	5	-.43	.27	-.34	.54
PF8	Picks up new learning quickly and without the need for repeated explanation	3.55	4.00	.92	2	5	-.10	.27	-.78	.54
PF9	Takes the time to understand the needs of Māori	3.37	3.00	.96	1	5	-.17	.27	-.25	.54
PF10	Proactively integrates their understanding of Māori needs when relating to Māori	3.40	3.00	1.04	1	5	-.15	.27	-.36	.54
PF11	Can bounce back quickly from setbacks	3.44	3.50	1.03	1	5	-.23	.27	-.58	.54
PF12	Thinks ahead, plans in advance	3.41	4.00	1.16	1	5	-.35	.27	-.83	.54
PF13	Explores alternative options prior to taking action	3.35	3.00	1.08	1	5	-.42	.27	-.35	.54
PF14	Is a trustworthy individual; is a person of their word	4.03	4.00	.99	1	5	-1.20	.27	1.22	.54
PF15	Pays attention to the environment and situational cues	3.53	4.00	1.03	1	5	-.55	.27	-.16	.54
PF16	Is in tune with his/her own feelings, moods, and emotions	3.63	4.00	.90	2	5	-.29	.27	-.61	.54
PF17	Expresses feelings to the right degree, to the right people and at the right time	3.45	3.00	.96	1	5	-.12	.27	-.56	.54
PF18	Can tune in to what makes people tick and connect with others	3.33	3.00	.98	1	5	-.29	.27	-.16	.54
PF19	Will consider others' feelings when decision making	3.33	3.00	1.05	1	5	-.23	.27	-.62	.54
PF20	Can effectively manage their own emotions, both positive and negative	3.53	4.00	1.02	1	5	-.45	.27	-.40	.54
PF21	Can positively lift others' mood, feelings, and emotions; can lift energy	3.37	3.00	1.05	1	5	-.38	.27	-.30	.54
PF22	Can remain calm and focused during periods of high stress and pressure	3.51	4.00	1.08	1	5	-.58	.27	-.23	.54
PF23	Is considered by colleagues to be a team player demonstrating loyalty to the team	3.74	4.00	1.10	1	5	-.80	.27	.16	.54

Table A3

Pattern Coefficients (Loadings) for GRT2 Factors 1 to 3

	Factors				Factors (cont.)				Factors (cont.)		
	1	2	3		1	2	3		1	2	3
AR2 Item 16	.59			AR2 Item 21	.31			AR2 Item 9			
NR2 Item 4	.58			AR2 Item 1				VR2 Item 27			
NR2 Item 16	.57	.34		AR2 Item 14				NR2 Item 9			
NR2 Item 7	.55			AR2 Item 3				VR2 Item 31			.87
AR2 Item 10	.53			NR2 Item 6				VR2 Item 34			.79
NR2 Item 11	.53			AR2 Item 22				VR2 Item 33			.70
NR2 Item 14	.52			NR2 Item 10				VR2 Item 29			.66
AR2 Item 19	.52			AR2 Item 15				VR2 Item 30			.62
NR2 Item 2	.52			VR2 Item 15				VR2 Item 35			.43
NR2 Item 13	.52			VR2 Item 2				VR2 Item 28			.35
AR2 Item 2	.51			VR2 Item 25				NR2 Item 5			.30
NR2 Item 17	.46	.31		VR2 Item 6				AR2 Item 20			
AR2 Item 17	.44			AR2 Item 5				NR2 Item 3			
VR2 Item 17	.42			VR2 Item 14				VR2 Item 26			
VR2 Item 10	.41			VR2 Item 8				VR2 Item 9			
NR2 Item 1	.41			AR2 Item 6				NR2 Item 12			
AR2 Item 12	.41			VR2 Item 19				VR2 Item 11			
VR2 Item 23	.39			VR2 Item 18				VR2 Item 32			
AR2 Item 8	.38			NR2 Item 20	.36	.55		VR2 Item 7			
AR2 Item 11	.38			NR2 Item 24		.54		VR2 Item 22			
AR2 Item 13	.38			NR2 Item 25		.51		VR2 Item 24			
VR2 Item 5	.37			NR2 Item 22		.50		AR2 Item 4			
AR2 Item 25	.36			NR2 Item 21		.48		VR2 Item 13			
AR2 Item 7	.35			NR2 Item 23		.47		VR2 Item 20			
AR2 Item 24	.34			NR2 Item 18		.46		VR2 Item 3			
NR2 Item 15	.33			NR2 Item 19		.42					
NR2 Item 8	.32			VR2 Item 1		-.39					
AR2 Item 23	.31			VR2 Item 16		-.38					
AR2 Item 18	.31			VR2 Item 12		-.33					
VR2 Item 21	.31			VR2 Item 4		-.32					

Note.
Factor loadings < .3 are suppressed.
VR2, NR2, and AR2 refer to the verbal, numerical, and abstract reasoning subscales of the GRT2.

Table A4

Structure Coefficients (Loadings) for GRT2 Factors 1 to 3

	Factors				Factors (cont.)				Factors (cont.)		
	1	2	3		1	2	3		1	2	3
AR2 Item 16	.62			AR2 Item 21	.32			AR2 Item 9			
AR2 Item 19	.57			AR2 Item 22	.32			VR2 Item 27			
NR2 Item 16	.55	.30		NR2 Item 8	.31			NR2 Item 9			
AR2 Item 10	.55			VR2 Item 15	.31			VR2 Item 31			.86
AR2 Item 2	.54			NR2 Item 6	.31			VR2 Item 34			.79
NR2 Item 2	.53			NR2 Item 10	.31			VR2 Item 33			.70
NR2 Item 4	.51			AR2 Item 1				VR2 Item 30			.64
NR2 Item 7	.51			VR2 Item 2				VR2 Item 29			.62
NR2 Item 14	.51			VR2 Item 6				VR2 Item 35			.47
NR2 Item 13	.50			AR2 Item 3				NR2 Item 5	.30		.36
NR2 Item 11	.50			AR2 Item 14				VR2 Item 28			.36
VR2 Item 17	.47		.31	VR2 Item 25				AR2 Item 20	.32		.33
AR2 Item 17	.46			VR2 Item 14				NR2 Item 3			
NR2 Item 17	.44			AR2 Item 5				VR2 Item 26			
AR2 Item 11	.43			AR2 Item 6				NR2 Item 12			
AR2 Item 12	.42			VR2 Item 8				VR2 Item 9			
VR2 Item 23	.42			VR2 Item 18				VR2 Item 11			
NR2 Item 1	.40			VR2 Item 19				VR2 Item 32			
AR2 Item 25	.40			NR2 Item 24		.54		VR2 Item 22			
AR2 Item 13	.39			NR2 Item 20	.34	.53		VR2 Item 24			
VR2 Item 5	.38			NR2 Item 25		.51		VR2 Item 7			
VR2 Item 10	.37			NR2 Item 22		.49		VR2 Item 13			
AR2 Item 8	.37			NR2 Item 21		.49		AR2 Item 4			
AR2 Item 7	.37	-.32		NR2 Item 23		.46		VR2 Item 20			
AR2 Item 24	.36			NR2 Item 18		.45		VR2 Item 3			
NR2 Item 15	.34			NR2 Item 19		.40					
AR2 Item 23	.34			VR2 Item 1		-.39					
VR2 Item 21	.34			VR2 Item 16		-.38					
AR2 Item 15	.33			VR2 Item 12	.32	-.35					
AR2 Item 18	.33			VR2 Item 4		-.32					

Note.
Factor loadings < .3 are suppressed.

Table A5

Reliability Analysis of Items in the Verbal Reasoning Subscale

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
VR2 Item 1	18.87	24.92	.24	.76
VR2 Item 2	18.98	24.55	.28	.75
VR2 Item 3	18.81	26.09	-.05	.77
VR2 Item 4	19.02	24.94	.18	.76
VR2 Item 5	18.90	24.76	.26	.75
VR2 Item 6	18.74	25.46	.20	.76
VR2 Item 7	19.56	26.00	-.02	.76
VR2 Item 8	18.86	25.71	.04	.76
VR2 Item 9	18.89	24.54	.32	.75
VR2 Item 10	19.15	24.95	.17	.76
VR2 Item 11	19.47	24.96	.22	.76
VR2 Item 12	19.02	24.53	.27	.75
VR2 Item 13	19.45	25.58	.07	.76
VR2 Item 14	19.07	24.50	.27	.75
VR2 Item 15	18.84	24.69	.32	.75
VR2 Item 16	19.16	24.37	.29	.75
VR2 Item 17	18.82	24.27	.47	.75
VR2 Item 18	19.03	25.18	.13	.76
VR2 Item 19	19.50	25.40	.13	.76
VR2 Item 20	18.85	25.87	.00	.77
VR2 Item 21	18.92	24.37	.34	.75
VR2 Item 22	19.11	24.22	.32	.75
VR2 Item 23	19.28	24.34	.30	.75
VR2 Item 24	18.92	25.09	.18	.76
VR2 Item 25	19.11	25.11	.14	.76
VR2 Item 26	19.44	25.03	.20	.76
VR2 Item 27	19.26	25.39	.08	.76
VR2 Item 28	19.15	23.94	.38	.75
VR2 Item 29	18.97	24.15	.37	.75
VR2 Item 30	19.22	23.32	.51	.74
VR2 Item 31	19.09	23.14	.56	.74
VR2 Item 32	19.49	25.08	.21	.76
VR2 Item 33	19.30	23.70	.44	.74
VR2 Item 34	19.09	23.53	.47	.74
VR2 Item 35	19.45	24.43	.35	.75

Note.

α = .76, N = 88.

Table A6

Reliability Analysis of Items in the Numerical Reasoning Subscale

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
NR2 Item 1	11.47	22.57	.33	.81
NR2 Item 2	11.80	21.66	.53	.80
NR2 Item 3	11.61	23.46	.11	.82
NR2 Item 4	11.40	22.54	.37	.81
NR2 Item 5	11.57	22.32	.36	.81
NR2 Item 6	11.49	22.41	.36	.81
NR2 Item 7	11.53	22.00	.44	.81
NR2 Item 8	11.34	23.26	.22	.82
NR2 Item 9	11.42	23.56	.12	.82
NR2 Item 10	11.58	22.38	.35	.81
NR2 Item 11	11.58	21.81	.47	.81
NR2 Item 12	11.45	23.31	.17	.82
NR2 Item 13	11.59	21.65	.51	.80
NR2 Item 14	11.39	22.49	.39	.81
NR2 Item 15	11.77	22.22	.39	.81
NR2 Item 16	11.70	21.25	.60	.80
NR2 Item 17	11.73	21.58	.53	.80
NR2 Item 18	11.75	22.76	.27	.82
NR2 Item 19	11.58	22.64	.29	.81
NR2 Item 20	11.98	22.18	.54	.80
NR2 Item 21	11.94	23.25	.22	.82
NR2 Item 22	11.85	22.50	.36	.81
NR2 Item 23	12.03	22.98	.39	.81
NR2 Item 24	11.97	23.39	.19	.82
NR2 Item 25	12.02	23.15	.31	.81

Note.
$\alpha = .82$, $N = 88$.

Table A7

Reliability Analysis of Items in the Abstract Reasoning Subscale

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
AR2 Item 1	14.07	24.04	.30	.81
AR2 Item 2	14.13	23.12	.49	.80
AR2 Item 3	13.97	24.63	.19	.81
AR2 Item 4	14.08	25.50	.00	.82
AR2 Item 5	14.11	24.26	.25	.81
AR2 Item 6	14.13	24.59	.18	.82
AR2 Item 7	13.70	24.65	.38	.81
AR2 Item 8	14.18	24.10	.29	.81
AR2 Item 9	14.22	24.36	.24	.81
AR2 Item 10	13.95	23.12	.53	.80
AR2 Item 11	14.31	23.64	.42	.81
AR2 Item 12	14.00	23.84	.35	.81
AR2 Item 13	13.82	24.06	.39	.81
AR2 Item 14	14.02	24.46	.22	.81
AR2 Item 15	13.80	24.35	.34	.81
AR2 Item 16	13.92	23.13	.55	.80
AR2 Item 17	14.22	23.39	.44	.80
AR2 Item 18	14.11	24.12	.28	.81
AR2 Item 19	14.14	22.90	.54	.80
AR2 Item 20	13.89	23.89	.39	.81
AR2 Item 21	13.91	23.88	.38	.81
AR2 Item 22	13.95	23.63	.41	.81
AR2 Item 23	14.14	23.54	.40	.81
AR2 Item 24	14.13	23.49	.41	.81
AR2 Item 25	14.13	23.26	.46	.80

Note.

α = .81, N = 88.

Table A8

Verbal Reasoning Item Difficulty and Discrimination Indices

	Item difficulty index (proportion correct)	Item discrimination index (item-total correlation)	Item discrimination index (corrected item-total correlation)
VR2 Item 1	.80	.31	.24
VR2 Item 2	.69	.36	.28
VR2 Item 3	.86	.02	-.05
VR2 Item 4	.65	.27	.18
VR2 Item 5	.77	.34	.26
VR2 Item 6	.93	.25	.20
VR2 Item 7	.11	.04	-.02
VR2 Item 8	.81	.12	.04
VR2 Item 9	.78	.40	.32
VR2 Item 10	.52	.26	.17
VR2 Item 11	.20	.30	.22
VR2 Item 12	.65	.36	.27
VR2 Item 13	.22	.15	.07
VR2 Item 14	.60	.35	.27
VR2 Item 15	.83	.39	.32
VR2 Item 16	.51	.38	.29
VR2 Item 17	.85	.52	.47
VR2 Item 18	.64	.22	.13
VR2 Item 19	.17	.20	.13
VR2 Item 20	.82	.08	.00
VR2 Item 21	.75	.42	.34
VR2 Item 22	.56	.41	.32
VR2 Item 23	.39	.39	.30
VR2 Item 24	.75	.26	.18
VR2 Item 25	.56	.23	.14
VR2 Item 26	.23	.28	.20
VR2 Item 27	.41	.18	.08
VR2 Item 28	.52	.46	.38
VR2 Item 29	.70	.45	.37
VR2 Item 30	.45	.58	.51
VR2 Item 31	.58	.62	.56
VR2 Item 32	.18	.28	.21
VR2 Item 33	.38	.52	.44
VR2 Item 34	.58	.54	.47
VR2 Item 35	.22	.42	.35

Table A9

Numerical Reasoning Item Difficulty and Discrimination Indices

	Item difficulty index (proportion correct)	Item discrimination index (item-total correlation)	Item discrimination index (corrected item-total correlation)
NR2 Item 1	.68	.41	.33
NR2 Item 2	.35	.60	.53
NR2 Item 3	.53	.21	.11
NR2 Item 4	.75	.45	.37
NR2 Item 5	.58	.45	.36
NR2 Item 6	.66	.44	.36
NR2 Item 7	.61	.52	.44
NR2 Item 8	.81	.30	.22
NR2 Item 9	.73	.21	.12
NR2 Item 10	.57	.43	.35
NR2 Item 11	.57	.55	.47
NR2 Item 12	.69	.26	.17
NR2 Item 13	.56	.58	.51
NR2 Item 14	.76	.46	.39
NR2 Item 15	.38	.48	.39
NR2 Item 16	.44	.66	.60
NR2 Item 17	.42	.60	.53
NR2 Item 18	.40	.36	.27
NR2 Item 19	.57	.38	.29
NR2 Item 20	.17	.60	.54
NR2 Item 21	.20	.30	.22
NR2 Item 22	.30	.44	.36
NR2 Item 23	.11	.44	.39
NR2 Item 24	.18	.27	.19
NR2 Item 25	.13	.37	.31

Table A10

Abstract Reasoning Item Difficulty and Discrimination Indices

	Item difficulty index (proportion correct)	Item discrimination index (item-total correlation)	Item discrimination index (corrected item-total correlation)
AR2 Item 1	.56	.39	.30
AR2 Item 2	.50	.57	.49
AR2 Item 3	.66	.28	.19
AR2 Item 4	.55	.10	.00
AR2 Item 5	.51	.34	.25
AR2 Item 6	.50	.28	.18
AR2 Item 7	.92	.43	.38
AR2 Item 8	.44	.38	.29
AR2 Item 9	.41	.33	.24
AR2 Item 10	.67	.60	.53
AR2 Item 11	.32	.49	.42
AR2 Item 12	.62	.44	.35
AR2 Item 13	.81	.46	.39
AR2 Item 14	.60	.31	.22
AR2 Item 15	.83	.40	.34
AR2 Item 16	.70	.61	.55
AR2 Item 17	.41	.52	.44
AR2 Item 18	.51	.37	.28
AR2 Item 19	.49	.61	.54
AR2 Item 20	.74	.46	.39
AR2 Item 21	.72	.45	.38
AR2 Item 22	.67	.49	.41
AR2 Item 23	.49	.48	.40
AR2 Item 24	.50	.49	.41
AR2 Item 25	.50	.54	.46

Table A11

Correlation Matrix for Performance Measure Items

	PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	PF10	PF11	PF12	PF13	PF14	PF15	PF16	PF17	PF18	PF19	PF20	PF21	PF22	PF23
PF1	1.00	.70	.50	.67	.63	.54	.56	.69	.55	.55	.64	.69	.75	.51	.67	.62	.63	.72	.58	.57	.66	.67	.57
PF2	.70	1.00	.72	.78	.63	.53	.64	.66	.51	.47	.63	.68	.69	.56	.68	.61	.55	.59	.62	.57	.63	.64	.73
PF3	.50	.72	1.00	.66	.47	.46	.52	.61	.57	.51	.49	.62	.57	.31	.57	.42	.43	.51	.48	.41	.47	.46	.50
PF4	.67	.78	.66	1.00	.65	.50	.77	.62	.60	.57	.76	.72	.69	.58	.69	.69	.64	.68	.75	.63	.73	.66	.81
PF5	.63	.63	.47	.65	1.00	.77	.65	.66	.55	.52	.65	.68	.69	.52	.72	.58	.62	.63	.65	.48	.56	.59	.65
PF6	.54	.53	.46	.50	.77	1.00	.53	.62	.48	.46	.47	.59	.55	.42	.53	.42	.39	.47	.48	.30	.42	.50	.52
PF7	.56	.64	.52	.77	.65	.53	1.00	.52	.61	.60	.76	.53	.60	.51	.59	.63	.65	.71	.77	.61	.76	.67	.73
PF8	.69	.66	.61	.62	.66	.62	.52	1.00	.44	.43	.61	.77	.73	.55	.75	.60	.61	.66	.52	.57	.61	.71	.63
PF9	.55	.51	.57	.60	.55	.48	.61	.44	1.00	.92	.63	.53	.53	.35	.49	.45	.47	.66	.64	.49	.63	.44	.43
PF10	.55	.47	.51	.57	.52	.46	.60	.43	.92	1.00	.65	.52	.56	.34	.47	.44	.50	.69	.65	.50	.64	.48	.44
PF11	.64	.63	.49	.76	.65	.47	.76	.61	.63	.65	1.00	.66	.67	.65	.63	.80	.73	.76	.78	.69	.84	.65	.71
PF12	.69	.68	.62	.72	.68	.59	.53	.77	.53	.52	.66	1.00	.87	.53	.78	.66	.68	.69	.69	.65	.65	.71	.66
PF13	.75	.69	.57	.69	.69	.55	.60	.73	.53	.56	.67	.87	1.00	.49	.79	.63	.70	.74	.69	.67	.72	.76	.66
PF14	.51	.56	.31	.58	.52	.42	.51	.55	.35	.34	.65	.53	.49	1.00	.58	.72	.65	.57	.54	.64	.59	.62	.79
PF15	.67	.68	.57	.69	.72	.53	.59	.75	.49	.47	.63	.78	.79	.58	1.00	.72	.73	.75	.65	.71	.71	.74	.68
PF16	.62	.61	.42	.69	.58	.42	.63	.60	.45	.44	.80	.66	.63	.72	.72	1.00	.78	.71	.70	.81	.73	.74	.72
PF17	.63	.55	.43	.64	.62	.39	.65	.61	.47	.50	.73	.68	.70	.65	.73	.78	1.00	.78	.75	.78	.72	.79	.68
PF18	.72	.59	.51	.68	.63	.47	.71	.66	.66	.69	.76	.69	.74	.57	.75	.71	.78	1.00	.75	.74	.78	.75	.66
PF19	.58	.62	.48	.75	.65	.48	.77	.52	.64	.65	.78	.69	.69	.54	.65	.70	.75	.75	1.00	.74	.74	.71	.72
PF20	.57	.57	.41	.63	.48	.30	.61	.57	.49	.50	.69	.65	.67	.64	.71	.81	.78	.74	.74	1.00	.75	.76	.68
PF21	.66	.63	.47	.73	.56	.42	.76	.61	.63	.64	.84	.65	.72	.59	.71	.73	.72	.78	.74	.75	1.00	.75	.73
PF22	.67	.64	.46	.66	.59	.50	.67	.71	.44	.48	.65	.71	.76	.62	.74	.74	.79	.75	.71	.76	.75	1.00	.75
PF23	.57	.73	.50	.81	.65	.52	.73	.63	.43	.44	.71	.66	.66	.79	.68	.72	.68	.66	.72	.68	.73	.75	1.00

Table A12

Loadings on Rotated Performance Factors 1 to 3

		Factor		
		1	2	3
PF20	Can effectively manage their own emotions, both positive and negative	.82		
PF16	Is in tune with his/her own feelings, moods, and emotions	.79		.31
PF17	Expresses feelings to the right degree, to the right people, and at the right time	.77	.32	
PF22	Can remain calm and focused during periods of high stress and pressure	.71	.45	
PF14	Is a trustworthy individual; is a person of their word	.65		
PF21	Can positively lift others' mood, feelings, and emotions; can lift energy	.65		.53
PF23	Considered by colleagues to be a team player demonstrating loyalty to the team	.62	.42	.40
PF18	Can tune in to what makes people tick and connect with others	.62	.40	.45
PF11	Can bounce back quickly from setbacks	.60		.59
PF8	Picks up new learning quickly and without the need for repeated explanation	.44	.74	
PF12	Thinks ahead, plans in advance	.50	.72	
PF13	Explores alternative options prior to taking action	.52	.66	
PF6	Communicates clearly with numbers and data		.65	.32
PF5	Communicates clearly orally and in writing	.33	.64	.39
PF15	Pays attention to the environment and situational cues	.60	.62	
PF2	Goes above and beyond the requirements of their role	.37	.61	.41
PF3	Has a strong desire to learn and invest in their personal development		.60	.40
PF7	Builds positive relationships, regardless of nationality, age, and gender	.43	.30	.72
PF9	Takes the time to understand the needs of Māori		.36	.64
PF19	Will consider others' feelings when decision making	.57	.31	.59
PF4	Voluntarily assists others, even when there is no personal benefit	.45	.49	.58

Note.
Factor loadings < .3 are suppressed.

Table A13

Reliability of Performance Appraisal; Item-Total Score Correlations

		Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
PF2	Goes above and beyond the requirements of their role	73.79	297.41	.78	.97
PF3	Has a strong desire to learn and invest in their personal development	73.55	301.81	.64	.97
PF4	Voluntarily assists others, even when there is no personal benefit	73.49	291.84	.85	.97
PF5	Communicates clearly orally and in writing	73.31	301.80	.77	.97
PF6	Communicates clearly with numbers and data	73.47	304.07	.61	.97
PF7	Builds positive relationships, regardless of nationality, age, and gender	73.28	299.48	.80	.97
PF8	Picks up new learning quickly and without the need for repeated explanation	73.44	301.68	.77	.97
PF9	Takes the time to understand the needs of Māori	73.62	303.69	.67	.97
PF10	Proactively integrates their understanding of Māori needs when relating to Māori	73.59	301.80	.67	.97
PF11	Can bounce back quickly from setbacks	73.55	296.12	.85	.97
PF12	Thinks ahead, plans in advance	73.58	292.98	.83	.97
PF13	Explores alternative options prior to taking action	73.64	294.88	.84	.97
PF14	Is a trustworthy individual; is a person of their word	72.96	302.56	.68	.97
PF15	Pays attention to the environment and situational cues	73.46	296.54	.83	.97
PF16	Is in tune with his/her own feelings, moods, and emotions	73.36	301.12	.81	.97
PF17	Expresses feelings to the right degree, to the right people, and at the right time	73.54	299.16	.81	.97
PF18	Can tune in to what makes people tick and connect with others	73.65	297.45	.85	.97
PF19	Will consider others' feelings when decision making	73.65	295.76	.83	.97
PF20	Can effectively manage their own emotions, both positive and negative	73.46	298.51	.78	.97
PF21	Can positively lift others' mood, feelings, and emotions; can lift energy	73.62	295.62	.84	.97
PF22	Can remain calm and focused during periods of high stress and pressure	73.47	295.32	.83	.97
PF23	Considered by colleagues to be a team player demonstrating loyalty to the team	73.24	294.68	.83	.97

Note.

N = 78. Performance item 1 (overall performance rating) is excluded.

Table A14

Descriptive Statistics for Performance Measure Items by Gender, and t-Test of Mean Differences

		M		*SD*		*t*-test for Equality of Means		
		Male	Female	Male	Female	*t*	*df*	Sig. (2-tailed)
PF1	Overall job performance	3.32	3.35	.91	1.02	-0.16	76	.87
PF2	Goes above and beyond the requirements of their role	3.17	3.23	1.05	1.09	-0.23	76	.82
PF3	Has a strong desire to learn and invest in their personal development	3.34	3.58	1.05	1.15	-0.95	76	.34
PF4	Voluntarily assists others, even when there is no personal benefit	3.43	3.61	1.16	1.20	-0.69	76	.49
PF5	Communicates clearly orally and in writing	3.55	3.87	.88	.96	-1.51	76	.14
PF6	Communicates clearly with numbers and data	3.43	3.65	1.02	1.05	-0.92	76	.36
PF7	Builds positive relationships, regardless of nationality, age, and gender	3.64	3.81	.97	.98	-0.75	76	.46
PF8	Picks up new learning quickly and without the need for repeated explanation	3.47	3.68	.83	1.05	-0.98	76	.33
PF9	Takes the time to understand the needs of Māori	3.28	3.52	.99	.89	-1.09	76	.28
PF10	Proactively integrates their understanding of Māori needs when relating to Māori	3.28	3.58	1.06	.99	-1.27	76	.21
PF11	Can bounce back quickly from setbacks	3.30	3.65	.93	1.14	-1.47	76	.15
PF12	Thinks ahead, plans in advance	3.30	3.58	1.12	1.21	-1.06	76	.29
PF13	Explores alternative options prior to taking action	3.28	3.45	1.04	1.15	-0.70	76	.49
PF14	Is a trustworthy individual; is a person of their word	4.06	3.97	.85	1.20	0.42	76	.68
PF15	Pays attention to the environment and situational cues	3.51	3.55	.98	1.12	-0.16	76	.88
PF16	Is in tune with his/her own feelings, moods, and emotions	3.60	3.68	.85	.98	-0.39	76	.70
PF17	Expresses feelings to the right degree, to the right people, and at the right time	3.40	3.52	.85	1.12	-0.50	76	.62
PF18	Can tune in to what makes people tick and connect with others	3.26	3.45	.97	1.00	-0.87	76	.39
PF19	Will consider others' feelings when decision making	3.21	3.52	1.06	1.03	-1.25	76	.22
PF20	Can effectively manage their own emotions, both positive and negative	3.57	3.45	.97	1.09	0.52	76	.60
PF21	Can positively lift others' mood, feelings, and emotions; can lift energy	3.30	3.48	.95	1.18	-0.77	76	.45
PF22	Can remain calm and focused during periods of high stress and pressure	3.49	3.55	1.00	1.21	-0.24	76	.82
PF23	Is considered by colleagues to be a team player demonstrating loyalty to the team	3.68	3.84	1.05	1.19	-0.62	76	.54

Note.

n = 47 (males), n = 31 (females).