Integrity, Conscientiousness, Neuroticism, and Ability: Relationships and Measurement.

A dissertation presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

at Massey University, Palmerston North,

New Zealand.

Paul Q. Wood

2011

Abstract

The purpose of this dissertation was to increase knowledge relevant to psychometrically oriented workplace selection and classification. Multivariate relationships among integrity, conscientiousness, neuroticism, and fluid and crystallised ability scales were investigated. Adverse impact and the capacity to use response time information as criteria of ability scoring were also investigated. These three foci all had the potential to contribute knowledge capable of increasing the accuracy of the measurement and interpretation of commonly used psychometric assessments.

Two cross-sectional studies were undertaken. The first study used archival data for extant assessments of ability, general personality, and integrity. It involved 211 participants having undertaken assessments as a function of job applications. The second study designed and piloted new scales of integrity, conscientiousness, neuroticism, and fluid and crystallised ability. It involved 317 participants who completed these scales online as voluntary participants.

The first study found integrity to be related to both conscientiousness and neuroticism, but not substantially related to ability. Conscientiousness was also negatively related to crystallised ability. These findings were replicated in the second study. The first study's neuroticism scale which included a suspicion/cynicism facet (i.e., subscale) had a negative relationship with ability indices. This finding was not replicated in the second study. This may have been due to the absence of a neuroticism facet measuring suspicion/cynicism in the second study.

Those identifying as Māori within the first study were found to score substantially less well than non-Māori on crystallised ability indices, but not other scales measured. Calculations suggested any resulting adverse impact could be reduced by combining ability assessments with scales of integrity, conscientiousness, and neuroticism. These calculations were based in part upon the assumption that relationships among assessments are likely to account for shared variance in job performance predictions. No significant differences were found in the second study; although the very small sample size used (N = 22) encourages caution regarding the robustness of this result.

Findings from the second study also suggested that relative to low-ability respondents, high-ability respondents took less time to complete crystallised items and more time to complete fluid ability items. A small significant relationship was also observed between conscientiousness and the length of time taken to complete the fluid ability scale.

The studies undertaken had a number of limitations. One limitation shared across these studies was the very small number of participants identifying as Māori (N46 in Study 1 and N22 in Study 2). Another common limitation was the inability to generalise findings based upon cross-sectional data drawn from participant groups of convenience rather than individuals selected via probability sampling.

Despite such limitations the preceding findings have a number of practical implications. One such implication is that relationships among scales may vary according to whether the level of analysis undertaken is at the Big Five or facet level and whose version of a scale is examined. On this basis practitioners should examine items in order to understand scale output, and researchers should examine relationships at the level of facet or ability subcomponent. Practitioners should also use personality assessments alongside those of ability if they wish to maximise predictive validity and reduce adverse impact for those identifying as Māori. Furthermore, the use of response time information in testing is probably better suited to controlling and checking respondents' approach to answering assessments than incorporation in scoring algorithms.

This dissertation makes two novel contributions concerning relationships between response time and participant characteristics. Firstly, negative relationships between ability indices and conscientiousness or neuroticism scales appear real. They do not appear to be a consequence of more conscientious or neurotic respondents taking longer to complete ability scales. Secondly, poor time-management strategies do not explain response time results that are inconsistent with the belief that higher-ability respondents will complete assessments more quickly than their lower-ability peers. Differences in the cognitive requirements associated with fluid and crystallised tasks instead appear to explain why higher-ability respondents take relatively less time to complete crystallised scales, but relatively more time to complete fluid ability scales.

Acknowledgements

Dante Alighieri's the Divine Comedy begins "*Mid way this way of life we're bound upon, I woke to find myself in a dark wood where the right road was wholly lost and gone.*" I once shared this fate…In much the same way as Virgil guided Dante, so too have countless others helped me find my "right road." My father Brian Wood has always being chief amongst such supporters. This dissertation provides him with another positive milestone along the journey of his most prodigal son. My brothers Jon, Andrew, and Chris have also played important parts. Jon and Chris always strongly encouraged me and the sibling rivalry associated with submitting before Andrew completed his Ph.D. at the London School of Economics provided additional motivation. My sister-in-law Cristiane also deserves a special mention for always enquiring after my progress and keeping me focused on my goals.

I will never be able to thank my supervisors adequately. There was always the unceasing support and advice, but their willingness to go beyond the call of duty leaves such platitudes better suited to other dissertations. Dr Gus Haberman and Dr Stephen Hill travelled hours out of their way to see me when I was constrained. They also made time to coax me through literacy lessons better suited to remedial English than Ph.D.s. Without their assistance I would not have submitted this dissertation.

The list of others who have supported me is long. There are many who have provided general encouragement or support, but the following list of people is limited to those who have made important contributions to the completion of this Doctorate: Dr Paul Englert, Dr Glenn Brown, Professor David Thomas (A.K.A. Uncle Dave), John Barlow, Mike Young, Henare Davidson, Susannah Sommerville, and Alison White. I would also like to say a special thanks to Chloé Hamman. She helped proof sections of this dissertation and served as a sounding-board for my ideas. As a function of this journey she has also tolerated a regularly stressed, distracted, and absent partner. I only wish my mother had lived to meet you. I dedicate this dissertation to my mother's memory:

Mary Jean Wood
1941 - 1995

Table of Contents

List of Tables and Figures

Note. The first numeral within Figures and Tables denotes the section of presentation and the second numeral denotes presentation order within section.

Chapter 1: Introduction

"*Hiring good people is hard. Hiring great people is brutally hard.*"

- Welch (2005, p.41).

Increased diversity, technological advances, globalisation, and operating within an uncertain economic climate ensure that the nature of employment continues to evolve. Positions within the workplace are becoming increasingly dynamic and less prescribed by specific job descriptions. This necessitates employees possessing more general competencies than was required in the past (Campion et al., 2011). Yet regardless of how the nature of employment may shift and change, the significance of employee performance remains unrivalled in its relevance to organisational effectiveness (Campion et al.; Cascio, 1991; Ilgen & Pulakos, 1991). As a result of this, a great deal of industrial, work, and organisational (I.W.O.) psychology's research continues to focus upon the job performance domain: job analysis, work design and motivation, training and development, and classification and selection (Ackerman & Humphreys, 1990; Campbell, 1999; Woods & West, 2010). This dissertation focuses on classification and selection.

Section 1.1: Dissertation Aims and Objectives

The aim of this dissertation is to enhance selection decisions by increasing the accuracy of psychometrically based classification through an increased understanding of how scales relate to each other, differ in average scores amongst relevant groups, and can be more accurately assessed. The use of psychometric assessments in selection decisions is ubiquitous and a lot is known about the relationships between these assessments and work related outcomes (e.g., Evers, Anderson, & Voskuijl, 2005; Grubb, Whetzel, & McDaniel, 2004; Schmidt & Hunter, 1998; Schmidt, Shaffer, & Oh, 2008; Thomas, 2004). Far less is known about the relationships between and among these assessments (Potosky, Bobko, & Roth, 2005; Reeve, Meyer, & Bonaccio, 2006; Wood & Englert, 2009). Section 1.2 outlines how such relationships have important implications for the optimal use and interpretation of psychometric results. Subsequent chapters within this dissertation will attempt to explore and clarify relationships between assessments of integrity, conscientiousness, neuroticism, and cognitive ability. This forms the first research objective of this dissertation.

The second research objective is to investigate whether there is a difference in personality or ability test scores between those identifying as Māori and non-Māori. Such differences can sometimes result in what is known as adverse impact. Adverse impact reflects the lower hiring ratios of ethnic minorities relative to majority group members (e.g., Campion et al., 2001; Grubb et al., 2004; Roth et al., 2001). This second research objective will also explore whether any such adverse impact can be reduced by combining ability with other assessments. Another key area for investigation concerns the meaning of psychometric scores themselves.

Psychometric assessments are only indirect indicators of constructs of interest. For example, a test of cognitive ability is not directly measuring one's cognitive ability. It is instead attempting to measure one's performance on a collection of tasks considered indicative of one's ability. How one performs on these tasks can be influenced by things outside of their actual ability. For this reason it is also important to examine whether or not any observed relationships are likely to be reflections of real-world interactions or consequences of test-taking approach/style. What research has examined relationships among assessment outcomes has provided little or no information in this regard. This is despite suggestions that personality differences may influence results through affecting test-taking style (e.g., Bates & Eysenck, 1993; Furnham, Forde, & Cotter, 1998; Zeidner, 1995). This suggestion raises a number of important questions for current and future explanations of observed relationships. For example, is the negative relationship between intelligence and conscientiousness a consequence of less intelligent people compensating through the adoption of more conscientious behaviour (Wood & Englert, 2009)? Or is it really due to relatively conscientious respondents taking longer to answer items? Some researchers have examined relationships among ability scores, personality traits, and response times on elementary cognitive tasks (Bates & Rock, 2004; Wickett & Vernon, 2000). Some have mooted the possibility that conscientious respondents are disadvantaged in timed standardised testing contexts (Jensen, 1998; Powers & Kaufman, 2004). One even examined the relationships among personality traits, response latencies when completing a personality assessment, and ability scores (Furnham, Forde, & Cotter, 1998a). Yet no one appears to have examined any interactions involving personality trait scores and the actual length of time taken to complete ability items.

Investigating relationships among personality traits, the length of time taken to answer ability items, and actual ability scores could provide an indication of whether or not relationships amongst personality traits and ability are a function of test-taking style. Yet an examination of tables of contents since 2000 for volumes of the *Journal of Educational Measurement, Applied Psychological Measurement, Psychometrika, Intelligence,* and *Personality and Individual Differences* failed to identify a single paper attempting to use response times on ability items to clarify observed relationships between personality traits and intelligence. Greater knowledge of the nature of such relationships will provide an advance in assessment interpretation and design. It also stands to influence the instructions provided to candidates completing ability assessments. Exploring such relationships among response time and respondent characteristics forms the third research objective of this dissertation.

A great deal has been written about nomological networks for individual predictors and common job performance criteria (e.g., Dalal, Lam, Weiss, Welch, Hulin, 2009; Gorman et al., 2011). Yet very little information is available about the nature of predictor intercorrelations amongst commonly used scales, and few if any assessments are interpreted or designed with knowledge of these relationships in mind (Moscoso & Iglesias, 2009; Pearce, 2009; Taylor, Keelty, & McDonnell, 2002). Nor do many of the commercially available assessments incorporate other advances in psychometrics (Clark, 2006; Borsboom, 2006a, 2006b). This dissertation will design and pilot a *composite* assessment of important psychometric predictors of future job performance. Within this dissertation, the term "composite" is intended to denote an assessment that is made up of distinct subtests (predictors) with the potential to facilitate the attainment of selection objectives (e.g., performance, fit, and diversity) more than any of its individual component parts. The predictors examined will be integrity, conscientiousness, neuroticism, and cognitive ability. Piloting a composite of these scales will facilitate the exploration of relationships between test-taking characteristics and predictor intercorrelations. It will also allow for the development of a composite assessment specifically intended to reduce the potential for adverse impact. This presents a first step towards answering the long standing call to more actively and seriously integrate findings from psychometrics into assessment test practice (Embretson, 1983, 1985; Mislevy, 1994). The research objectives and more general design aim of this dissertation can be stated as follows:

- ***Research Objective 1:*** To investigate whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. Furthermore, to explore variations among any such differences at the level of *facet* or ability *subcomponent*. (Within this dissertation facet and subcomponent refer to separately measurable parts, yet components of a broader scale or construct.)
- ***Research Objective 2:*** To investigate whether there is a difference in personality or ability test scores between those identifying as Māori and non-Māori. Then to explore whether any such differences could be reduced by combining ability tests with other assessments, which in part depends upon the extent to which variance in job performance is accounted for by shared variance amongst predictors (Sackett & Ellingson, 1997).
- ***Research Objective 3:*** To investigate whether the length of time taken to answer ability assessments is related to respondent ability or item difficulty. Moreover, to explore whether differences in conscientiousness or neuroticism scores meant respondents take longer to answer ability assessments.
- ***General Design Aim:*** Develop a composite assessment of examined predictors that will be able to incorporate predictor interactions and psychometric advances. The predictor interactions of relevance concern likely reductions in the adverse impact experienced by ethnic minorities and increases in predictive validity. Psychometric advances include the use of Item Response Theory for ability items and the collection of response time as a potential future scoring parameter.

Literature supporting the pursuit of these research objectives and design aim is provided in Chapter 2. The next section of this chapter will provide brief overviews of how pursuing these research objectives and aims may enhance the use of psychometrics amongst practitioners and the understanding of relevant constructs amongst theorists. Section 1.2.1 focuses on the potential practical value for practitioners.

Section 1.2: Importance of Study

Section 1.2.1: Practical Value

The accuracy of selection decisions impacts upon factors such as an organisation's bottom-line, culture, ethos, and reputation. This investigation has the potential to increase the accuracy of selection decisions in two ways. The first involves clarifying relationships amongst important predictors of job performance. Knowledge regarding the interrelationships among such predictors may facilitate the optimisation of calculations of future job performance based upon combining predictors, the detection of potentially counterproductive employees, and the selection of productive employees in terms of both task and contextual work behaviours. This stands to increase an employers' capacity to select the best potential employees, which is likely to reduce recruitment costs and improve productivity.

Multiple methods of collecting information intended to predict job performance can be utilised in the process of personnel selection. Common examples include: interviews, background checks, assessment centres, reference checks, biodata, work samples, integrity tests, personality tests, and intelligence and ability measures (Hunter & Hunter, 1984). Recent work within a New Zealand context has also focused on the capability of such measures to provide predictive information on competencies shown to impact objective work outcomes (Jackson, Cooper-Thomas, van Gelderen, & Davis, 2010). Decisions regarding the adoption of selection devices must depend upon how useful they will be to an organisation or employer. Does the device meet a documented business need, or add value *beyond* selection methods already in place (Wanek, 1999)? This latter requirement necessitates an understanding of the relationships between any assessments considered for addition to a selection system and those already use in that system. Understanding such relationships is useful for incremental validity, or maximising the overall criterion-related validity of a selection system. This is because the addition of a measure should increase criterion-related validity through accounting for previously unrealised aspects of the criterion. This makes the combination of relatively unrelated predictors that correlate highly with the criterion of interest highly desirable (Smith, Fischer, & Fister, 2003; Ghiselli, Campbell, Zedeck, 1981; Ones, 2001; Vernon, 1956).

Psychometrics are often used in combination (e.g., personality and cognitive ability), yet very little is known about their interactions (Chamorro-Premuzic & Furnham, 2004; Moutafi, Furnham, & Paltiel, 2005; Wood & Englert, 2009). Human resource professionals have tended to draw their own anecdotally based conclusions about the potentially compensatory or otherwise interactive nature of psychometric scores. The primary research objective within this dissertation is to explore interactions of commonly used psychometric predictors of performance. For example, does a high score on neuroticism increase the likelihood of counterproductive behaviour indicated by an integrity test score (Dalal et al., 2009)? Alternatively, does a high score on conscientiousness mean respondents with relatively less cognitive ability may compensate through greater diligence (Duckworth & Seligman, 2005; Wood & Englert, 2009) or vice versa (Postlethwaite, Robbins, Rickerson, & McKinniss, 2009; Schmidt, Shaffer, & Oh, 2008)? Clarifying the nature of such relationships may be advantageous to organisations wishing to standardise interpretation and increase the accuracy of selection decisions.

The literature review undertaken in Chapter 2 identifies cognitive ability, conscientiousness, neuroticism, and integrity as the relevant predictors within this investigation. Exploring relationships amongst integrity and other personality and ability assessments is also likely to increase the accuracy of exclusion or screening-out decisions made. Hiring employees who go on to display counterproductive workplace behaviours has serious negative consequences for the organisation (Connerley, Arvey, & Bernardy, 2001; Vardi & Weitz, 2004). Besides the costs in individual performance there are often more indirect negative consequences. One such example includes a reduction in the morale of other employees. This in turn results in the intentional withholding of effort and organisational-citizenship behaviours by those affected (Penney, Spector, & Fox, 2003). The seriousness and prevalence of such indirect effects is illustrated by the finding that employees' second most common workplace fear relates to negative coworker behaviour (Connerley et al., 2001; Ryan & Oestreich, 1991).

One way in which companies attempt to avoid hiring counterproductive employees is through background checks. According to Connerley et al. (2001) anywhere from 80-95% of North American corporations employ some kind of background check. Nearly all New Zealand organisations and recruitment firms surveyed use background/reference checks (Taylor, Keelty, & McDonnell, 2002). Large NZ corporations reported using reference checks 98% of the time for non-management positions, and 97% of the time for management positions. NZ recruitment firms reported using reference checks 100% of the time for management positions (Taylor et al., 2002). Yet such background checks are often only as good as they are thorough and largely depend on the inquisitorial skill of the inquirer. They are also very expensive and labour intensive (Hilliard, 2001). One way to increase the consistency of screening-out processes and reduce the cost and logistics of background checks is to use a standardised psychometric integrity test. A major controversy involving the use of integrity tests in workplace selection decisions is that such use may result in an individual being falsely labelled as dishonest through misclassification (Sackett & Wanek, 1996). This controversy has not prompted the discontinuation of integrity test use. Yet it has brought to the fore the importance of reducing misclassification rates where possible (see Sackett & Wanek, 1996 for review). Better understanding relationships between integrity and other difference measures may benefit potential employees by ensuring that their chances of being mistakenly excluded from a position in which they would have been competent are reduced. The identification of those personality and cognitive ability scores most strongly correlated with dishonesty scores may also facilitate a reduction in selecting people who present a higher risk of CWBs than indicated by an integrity test score alone.

A general developmental aim of this dissertation is to increase the accuracy of selection decisions by designing and piloting a cost-effective and efficient assessment that will have potential to incorporate predictor interactions and psychometric advances. Organisations are concerned about hiring the right people. The traditional psychometrics employed for this purpose appear to focus upon the constructs measured rather than their implications for an individual's fit with role and organisational requirements. For example, putting candidates through personality assessments more concerned with content validity than whether or not all the traits assessed are of relevance to job performance. While many such assessments have criterion-validity studies available, many of those involved in selection decisions do not reference this information and instead hypothesise which personality preferences are relevant. Such hypotheses are likely to be influenced by self-evaluation biases leading to favouritism of those displaying trait or social-identity similarity to that of the rater (Higgins & Scholer, 2008). Selectors are also unlikely to consistently and proportionately weight the traits considered. The potential impact of this will be reduced via the inclusion of only factors shown to be consistently predictive of job performance. It will also be reduced through a combination of strong and generic indicators of job performance rather than factor scales intended to maximise coverage of a construct domain.

A psychometrically sound composite of predictors is also likely to provide organisations with greater predictive validity then many of the other assessments available in the market (Clark, 2006; Borsboom, 2006a, 2006b). The parsimony of this composite is also likely to reduce the recruitment costs incurred by organisations and the stresses experienced by job applicants. The ability to reduce the time required for the psychometric portion of workplace selection is expected to lessen both applicant stress and inconvenience. The ability of a recruiting organisation's personnel to administer a single assessment rather than a collection of assessments will also reduce the resources usually devoted to this function. This reduction includes the additional cost of having to use multiple psychometric tests to achieve a comparable result.

It is also hoped that the composite assessment resulting from this dissertation will reduce the adverse impact experienced by some ethnic groups in selection decisions involving psychometrics. Significant mean differences in cognitive ability tests are observed in comparisons among some ethnic groups (Hunter & Hunter, 1984; Jensen, 1980, 1998; Loehlin, Lindzey, & Spuhler, 1975; Ones et al., 2005b; Roth, Bevier, Bobko, Switzer, & Tyler, 2001; Wittmann, 2005). Adverse impact occurs when this difference negatively affects the hiring ratios of lower scoring minority group members. As well as potentially impacting on the organisational performance of some companies, the presence of adverse impact can require an employer to engage in the time consuming and expensive process of defending the job-relatedness of their selection system (Dunleavy & Gutman, 2009; McDaniel, 2009). There are also a variety of serious societal costs associated with adverse impact, including the on-going wealth and educational disparities associated with employment opportunities (Herrstein & Murray, 1994).

Adverse impact has caused some researchers to question the ethics of using intelligence measures in selection decisions (Blits & Gottfredson, 1990; Boehm, 1972; Gael & Grant, 1972; Gould, 1993; Humphreys, 1973; Ones et al., 2005b; Schmidt, Berner & Hunter, 1973; Schmidt & Hunter, 1974). Despite some job complexity associated differences in the predictive validity of ability assessments, it is also argued that the predictive validity of cognitive ability tests is so high that other selection methods should only be considered in terms of the incremental validity they can add to intelligence-based selection decisions (Grubb et al., 2004; Ones & Viswesvaran, 2001b; Ree, Earles & Teachout, 1994; Schmidt & Hunter, 1998). Research suggests that New Zealand organisations share this confidence in the predictive power of cognitive ability measures (Ryan, McFarland, Baron & Page, 1999; Taylor et al., 2002). Unfortunately research has also shown New Zealand may share the negative impact of cognitive ability measures upon minority racial groups (Beck & St George, 1983; Guenole, Englert & Taylor, 2003; Harker, 1978; St George, 1983).

It has become an important goal for many organisations to reconcile the reduction of adverse impact with the maximisation of predictive validity (Cascio, Jacobs, & Silva, 2010; Cook, 2004; De Corte, Bobko, & Roth, 2008; De Corte, Lievens, & Sackett, 2008; Ones, Viswesvaran, & Dilchert, 2005a; Pulakos & Schmitt, 1996). This goal is particularly relevant in New Zealand with commitments to equal opportunity employment under the State Sector Act 1988 and Treaty of Waitangi. One effective way to minimise adverse impact is to use cognitive ability measures in combination with performance predictors that do not negatively impact upon minorities (De Corte et al., 2008; Potosky et al., 2005, 2008; Sackett, De Corte, & Lievens, 2008). The addition of other relevant unbiased measures lessens the impact of cognitive ability scores by providing other performance relevant information worthy of consideration and by accounting for shared variance in job performance predictions (Bobko, Roth, & Buster, 2007; Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). When this dissertation refers to a test as "biased" it suggests the test produces different average scores for different groups. The term does not denote support for any particular position concerning the origins of such differences. The measures of integrity and personality employed within the investigation both assess constructs that have been shown to be predictive of future job performance without displaying significant ethnicity related differences (Cullen & Sackett, 2004; Evers, Te Nijenhuis, & Van Der Flier, 2005; Foldes, Duehr, & Ones, 2008; McCrae et al., 1999; Ones, 1993; Ones & Viswesvaran, 1998a, 1998b, 1998c; Sackett, 1994). Combining ability, integrity, and personality in the composite developed is expected to result in both a considerable reduction in adverse impact and an increase in predictive validity through a attenuation of bias (Evers et al., 2005). Measuring facets of ability and personality is also expected to contribute to a reduction in adverse impact. This is based on findings that group differences responsible for adverse impact are less prominent for specific abilities and traits than more general cognitive abilities or big five level traits (Hough, Oswald, & Polyhart, 2001; Sackett & Lievens, 2008).

Section 1.2.2: Contribution to Knowledge

This dissertation contributes to the body of knowledge within the fields of personnel selection and differential psychology. The direction of future research within the domain of job performance and measurement will be significantly affected by research examining predictors of job performance (Arvey & Murphy, 1998; Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Wood & Englert, 2009). In their review of personnel selection Borman et al. (1997) strongly urged further research to identify and verify relationships between job performance and individual differences such as integrity, personality, and ability. Furthermore, a better understanding of the integrity construct may be gained by this dissertation's systematic exploration of relationships between integrity tests and other measures in the temperament and cognitive domains. As well as contributing to an understanding of the integrity construct, this dissertation may also address a number of methodological concerns.

Investigators of integrity testing have appealed for more independent research in this area (Camara & Schneider, 1994; Ones & Viswesvaran, 1998b). This dissertation includes an independent examination of the widely used Stanton Survey of Integrity (SSI). Many investigations into integrity tests have also suffered from a restricted range due to their focus upon incumbents instead of applicants (Borman & Motowidlo, 1997). The participant group detailed in section 4.3.1 is comprised of New Zealanders within a job application context. This also addresses the limitation for previous integrity test researches relative lack of data collection outside of the U.S.A. (Ones & Viswesvaran, 2001a; Wanek, 1999) and the overuse of non-applicant participant groups (Landy, Shankster & Kohler, 1994).

As well as addressing the above issues concerning research methodology, this dissertation may help further clarify the constructs of integrity and Big Five conscientiousness and neuroticism. Sackett and Wanek (1996) conducted a review of over 200 journal articles, books, convention papers, and other sources before drawing the conclusion that the construct of integrity is ill defined. Considerable work has since helped clarify the integrity construct through examining its relationships with Big Five factors of personality (Wanek, Sackett, & Ones, 2003). Yet relatively little is known about the relationship between integrity and the facets from which the Big Five are derived. This is an area needing further investigation (Berry, Sackett, & Wiemann, 2007; Cullen & Sackett, 2004). Among other things, the first study undertaken within this dissertation will explore relationships between integrity and Big Five neuroticism facets of affected by feelings, suspicious, self-doubting, and tense-driven. The first study will also explore relationships between integrity and Big Five conscientiousness facets of restrained, self-disciplined, and conscientious.

Exploring relationships at the facet-level is also a particularly apposite line of inquiry considering argument over the predictive merits of Big Five versus its facet components within an applied environment (Boyle, Stankov, & Cattell, 1995; Hough & Oswald, 2000; Judge, Klinger, Simon, & Wen Fen Yang, 2008; Kline, 1995; Mershon & Gorsuch, 1988; Murphy & Dzieweczynski, 2005; Pearce, 2009; Wood & Englert, 2009). The legitimacy of such concerns is exemplified by findings suggesting that Big Five facets are better predictors than the Big Five factors they comprise. For instance, Ashton, Jackson, Paunonen, Helmes, and Rothstein (1995) discovered that the Big Five extroversion facets of affiliation and exhibition are better at predicting fun-seeking behaviour than overall extroversion itself. This same study also found that the facet of dominance predicted surgent behaviour, such as speaking out in class, better than the global extroversion factor. The meta-analysis of Dudley, Orvis, Lebiecki, and Cortina (2006) had a similar finding. Dudley et al. found conscientious facets of achievement, dependability, order, and cautiousness incrementally predict performance above and beyond the superordinate-level conscientiousness trait. Ashton (1998) reported similar results for workplace delinquency. Ashton found the facet scales of risk-taking and responsibility to be better predictors of delinquency than the conscientiousness superordinate factor from which they were derived.

The examination of relationships between integrity scores and facet-level personality traits undertaken in section 4.4 may help clarify the integrity construct and contribute to the superordinate versus facet trait debate. This facet-level of analysis also complies with well documented advantages concerning construct clarity and incremental validity (Smith et al., 2003). Furthermore, item writing undertaken for the Big Five constructs of conscientiousness and neuroticism detailed in section 5.2.2 covers the content domains of competing facet-level models. Because of disagreements concerning the content and domain overlap between the predictors examined, the factor structure of these constructs determined in section 5.3 uses exploratory factor analysis rather than any *a priori* models.

This dissertation may also contribute to our understanding of the relationship between integrity and intelligence by differentiating between fluid and crystallised intelligence. No studies making such a differentiation were discovered during a substantial literature search. Making this differentiation has previously clarified relationships between other personality-based constructs and intelligence (Moutafi, Furnham & Crump, 2003; Moutafi, Furnham & Paltiel, 2004; Wood, 2004: Wood & Englert, 2009). The clarification achieved via this differentiation may assist in subsequent model and theory development. It is also likely to promote hypothesis generation and provide impetus for future investigations.

The composite assessment designed contributes to a personnel selection approach attempting to neutralise the potentially adverse impact of cognitive ability tests through exploring their use in combination with putatively race-neutral integrity and personality tests (Campbell, 1996; De Corte et al., 2008; Ones & Viswesvaran, 1998a). In doing so this dissertation may contribute to a reduction in adverse impact on the same minority group members some believe to have historically suffered at the hands of psychometricians (Gould, 1993; Richardson, 1991, 1998; Richardson & Spears, 1972).

Section 1.3: Chapter Overviews

Chapter 2: Literature Review (What we know)

Chapter two explores the recruitment context, and which psychometric predictors of performance should be examined on the basis of individual predictive validity, incremental validity, performance domain coverage, and adverse impact reduction. It then details how these predictors have previously been defined and used in the workplace, and what is known of their relationships. Chapter two then looks at response time within the context of response styles, test and respondent characteristics, and as a potential explanation of previously observed relationships between personality trait and ability test scores.

Chapter 3: Research Objectives

Chapter 3 provides and brief summary of the three research objectives and their rationale. These are based upon information presented in the preceding sections and Chapter 2.

Chapter 4: Study 1 (Predictor Relationships and Differences).

Chapter 4 reports the method, results, and discussion for Study 1. The method section details demographic characteristics of the participants within the archival data employed, the measures and procedures that were used, and the rationale for the analyses chosen. The results section investigates the first two research objectives of this dissertation. The discussion section details the largely replicated relationships among integrity, conscientiousness, neuroticism, and fluid and crystallised ability. It examines the implications of these findings and explores failures to replicate previous findings. The discussion then examines differences in predictor scores between those identifying as Māori and non-Māori. The extent to which observed differences in ability indices are likely to adversely affect Māori is then explored, as is the extent to which combining assessments may reduce this impact. Chapter 4 also explores specific limitations and potential directions for future research.

Chapter 5: Study 2 (Composite Design and Relationship Clarification).

Chapter 5 reports Study 2's method, results, and discussion. The method section details the demographic characteristics of participants, the measures designed, the procedures that were used, and the rationale for the analyses chosen. The results section details the processes of item selection, and tentative test validation and reliability. It then explores the three research objectives previously outlined. The discussion section reports on very similar findings to Study 1 regarding the first two research objectives. It then comprehensively addresses outcomes relevant to the use of response time in the administration and scoring of assessments. The discussion section also describes the small impact of conscientiousness and one neuroticism facet on the length of time taken to answer ability assessments. It also reports on the composite designed and specific limitations and potential directions for future research.

Chapter 6: Integration and Conclusions.

Chapter 6 integrates the findings of Study 1 and Study 2. It looks at how well the dissertation achieved its research objectives and explores discrepancies between results for Study 1 and Study 2. It then discusses general limitations, directions for future research, and themes that emerged.

Chapter 2: Literature Review:

Recruitment and Psychometric Prediction

This chapter provides the context for research objectives and assessment development. A review of selection and classification literature quickly identifies psychometric assessments of integrity, personality, and intelligence as of primary interest to the goals of this dissertation (Thomas, 2004). Each of these constructs will be examined separately in subsequent sections. The following analyses are not intended to be definitive. More thorough analyses are beyond the scope of this dissertation and are available elsewhere. Sections 2.2 to 2.4 provide cumulative information on relationships among the predictors examined. Sections providing information on maximising validity, reducing adverse impact, and response time are also included. Section 2.1 below expands upon the context and rationale for predictor choice.

Section 2.1: Context and Rationale for Predictor Choice

This section outlines predictor choice considerations. This is done via subsections concerning the recruitment context, the individual utility and ubiquity of various predictors, the combined utility of predictors, and the potential for different combinations of predictors to result in a reduction in adverse impact.

Section 2.1.1: The Recruitment Context

The purpose of this dissertation is to increase the accuracy of selection decisions by increasing knowledge of the construct domains and measurement on which psychometrically oriented classification and selection depend. Campbell, Gasser, and Oswald (1996) demonstrated a large disparity between high and low achiever performance in both high and low complexity jobs. Campbell et al. found performance ratios amongst employees to range from 2:1 to 4:1 for low complexity jobs and between 6:1 and 10:1 for jobs of greater complexity. This disparity remained consistent when definitions of performance were altered, and when those likely to have been the worst performers had already been excluded via selection procedures. Job performance in the workplace can be optimised via either or both of two general procedures: improving individuals for the better once they are hired, or improving the process for selecting employees in the first place (Campbell et al., 1996). This dissertation focuses on improving the selection process.

The current economic climate makes ensuring recruitment procedures actually result in the selection of productive employees essential. The price of selecting, hiring, and terminating employees is high. So is the price of losing superior workers to competitors (Cook, 2004). As evidenced in Campbell et al.'s (1996) ratio differences, the workplace is comprised of employees of widely varying degrees of proficiency. Mistakes in the selection process can result in greater absenteeism and turnover, reduced productivity, decreased morale, and even monetary loss as a consequence of lacklustre performance, theft, or fraud (Cascio, 1991). According to Dunn (1995) the average cost of increased turnover is approximately one and a half times greater than the annual salary of the vacated positions. The cost to organisations of counterproductive workers is thought to run into billions of dollars in both direct and indirect costs (Billings, 2001; Murphy & Lee, 1994b; Penney, Spector, & Fox, 2003). Good workers do twice as much as poor workers, and the difference in value between a good and a poor worker is roughly equal to the salary they are paid (Cook, 2004). Moreover, as an unintended consequence of legislation designed to protect employee rights the dismissal of substandard employees can be expensive, lengthy, and legally problematic (New Zealand Press Association, 2005). For these reasons it is advantageous for organisations to consistently and effectively select relatively productive employees. This dissertation seeks to provide knowledge concerning which measures and combinations thereof may be designed and/or used by organisations to enhance the prediction of successful job performance.

Section 2.1.2: Predictor Choice

The last half of the twentieth century saw considerable research aiming to determine an organisation's capacity to predict a potential employee's future job performance (Schmidt & Hunter, 1998). Consistent among such research has been the finding that cognitive ability is the single best predictor of job performance and training success (Evers et al., 2005; Grubb, Whetzel, & McDaniel, 2004;

Schmidt, Shaffer, & Oh, 2008). Cognitive abilities are characteristics such as verbal or numerical reasoning ability that vary among individuals or groups of individuals (Corsini & Auerbach, 1998). The term cognitive ability is often used within psychometrics as a synonym of intelligence as measured by tests. The terms cognitive ability and intelligence are also used interchangeably within this dissertation and refer to the construct domain covered by tests of cognitive ability/intelligence. Unless otherwise specified, this dissertation also uses "ability" as shorthand for cognitive ability. An explication of cognitive ability theories, models, and definitions is contained within section 2.4 of this chapter.

Meta-analytic findings suggest the predictive validity of ability tests vary only marginally across job types (Hunter & Hunter, 1984; Ones, Viswesvaran, & Dilchert, 2005a, 2005b; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Hunter, & Pearlman, 1981). They also suggest ability remains the best predictor of job performance across levels of job complexity (Grubb et al., 2004; Hunter, 1980) and cultures (Ones et al., 2005a, 2005b). The utility of ability in predicting employee training efficiency has also received substantial support (Bertua, Anderson, Salgado, 2005; Hunter, 1986; Hunter & Hunter, 1984; Ones et al., 2005a, 2005b; Ree & Earles, 1992). Furthermore, ability continues to predict job performance long after substantial on-the-job experience has been acquired (Gottfredson, 1997; Hunt, 1995a, 1995b; Ones et al., 2005b). Although these findings fail to take into account limitations such as Jensen's (2002) law of diminishing returns for cognitive ability and job performance, the do legitimately support the inclusion of cognitive ability amongst predictors examined.

Personality assessments also appear worthy of examination (Kanfer, Wolf, Kantrowitz, & Ackerman, 2010). The applied use of personality measures in the workplace continues to grow (Murphy, 1996; Salgado, 1999). New Zealand recruitment firms increased their use of personality assessments from 67% to 89% within the space of a decade (Taylor et al., 2002). There are good reasons for this rapid increase in prevalence. Meta-analytic research suggests that between 10% and 30% of variance in job performance is explicable via personality trait differences (Furnham, 2001). An individual's choice of work, reaction to it, and productivity therein are all influenced by their personality traits (Lowman, 1996). Ones success in their chosen vocation, potential for burnout, trainability, and subsequent job satisfaction are all influenced by their personality makeup (Cropanzano, James, & Konovsky, 1993; George, 1996; Maslach, Schaufeli, & Leiter, 2001; Spencer & Spencer, 1993). Moreover, the personality traits of conscientiousness and neuroticism consistently predict job performance better than other personality traits (Barrick, Mount, & Judge, 2001; Barrick & Zimmerman, 2009; Evers et al., 2005).

Investigations supporting the validity of integrity tests have also continued to suggest applicants who score poorly on integrity tests make worse employees (Berry et al., 2007; Murphy, 1993; Ones, Viswesvaran & Schmidt, 1991; Sackett & Wanek, 1996). Schmidt and Hunter's (1998) investigation into the validity and utility of selection methods also suggest personality traits and integrity measures increase incremental validity when added to ability assessments. Schmidt and Hunter's meta-analysis suggested that integrity and conscientiousness provide a 27% and 18% increase respectively to the validity accounted for by ability alone. Of 18 predictors integrity and conscientiousness provided the greatest incremental increase among psychometric measures. Subsequent investigations examining integrity test capacity to add incremental validity to ability predictions provide even more support (Marcus, Wagner, Poole, Powell, & Carswell, 2009).

Ability and personality assessments seem likely to account for unique variance in job performance. Incremental and individual validity considerations are further discussed in section 2.1.4. Yet before examining these predictors in more detail, a brief inspection of what is meant by "job performance" is in order.

Section 2.1.3: The Job Performance Domain

Research over the past two decades has made considerable inroads into understanding and explicating the construct of job performance and its interrelationships with predictors (Chan, 2005). Although consensus remains elusive, a common pattern among job performance models is emerging. Numerous theorists are proposing job performance models comprising two distinct qualitative categories necessary for organisational effectiveness: job task performance and job contextual performance

(Arvey & Murphy, 1998; Chan, 2005; Stone-Romero, Alvarez, Thompson, 2009). Job task performance concerns those core work behaviours that directly pertain to a person's job functions concerning an organisation's primary objectives. Task performance is generally predicted via intelligence or ability tests (Bacha, 2003; Chan, 2005; Collins & Griffin, 1998; LePine & Van Dyne, 2001; McHenry, Hough, Toquan, Hanson, & Ashworth, 1990; Penney & Borman, 2005; Wise, McHenry, & Campbell, 1990). Job contextual performance concerns "extrarole" or peripheral behaviours that are not always explicitly required within a role. Although often not explicitly part of a job, contextual behaviours impact upon an organisation's profitability, working environment, and ethos. This makes the inclusion of contextual performance within the criterion domain of job performance essential for organisational effectiveness (Penney & Borman, 2005). According to Penney and Borman (2005) contextual performance involves activities like volunteering for additional work, helping and being co-operative, following rules and procedures, and persisting with extra effort and enthusiasism. Other related examples of behaviours associated with contextual performance include prosocial/citizenship behaviour, stealing, tardiness, and absenteeism. Contextual performance can be predicted via personality and integrity tests (Chan, 2005; Collins & Griffin, 1998; Hilliard, 2001; Hurtz & Donovan, 2000; Kaufman & Borman, 2004; LePine & Van Dyne, 2001; McHenry et al., 1990; Osborn & Marion, 2009; Penney & Borman, 2005; Wise et al., 1990).

The importance of task performance to organisational efficiency has long been recognised. Knowledge of contextual performance's role in organisational success is now developing too. This has seen the release of a variety of reviews and publications emphasising non-task specific components of contextual performance, such as *trust* (Ciancutti & Steding, 2001; Jones, Couch, & Scott, 1997; Kramer & Cook, 2004; Lewis, 1999; Reina & Reina, 1999; West, 2002).

According to Reina and Reina (1999) there has never been a greater need for trust within the workplace than today. Ciancutti and Steding (2001) position trust alongside technology and innovation as one of the most powerful driving forces behind business success. The impact of trust upon "bottom-line" team and organisational performance is now well documented (Dirks & Daniel, 2004). It is also well documented that trust is more easily broken than established (Lewicki &

Wiethoff, 2000), and that the contextual component that integrity forms is crucial in the development and maintenance of trust (Baccili, 2001; Caron, 2003; Ciancutti & Steding, 2001; Norman, 2003). Consistent with this, supervisors consider both task and contextual aspects of performance equally important when evaluating subordinates (Borman, White & Dorsey, 1995; Motowidlo & Van Scotter, 1994). In fact, meta-analytic results suggest contextual and task performance explain 12% and 9.3% of variance in performance ratings respectively (Podsakoff, MacKenzie, Paine, & Bachrach, 2000). According to Podsakoff et al. (2000) contextual performance is absolutely crucial for the success of an organisation. Contextual performance enhances managerial and employee productivity; it facilitates work group functioning by improving coordination, communication, and relationships among group members; it frees up resources for more productive uses; it creates a more stable work environment; it improves an organisation's ability to recruit and retain better employees; and it allows for smoother transitions in times of change.

While such research supports the importance of contextual performance, organisations are unlikely to function if task performance is not also and already at an acceptable standard. The following section looks at how well the predictors chosen are likely to provide coverage of the job performance domain.

Section 2.1.4: Predictor Coverage of the Job Performance Domain

Figure 2.1 illustrates literature review based hypothesised relationships among ability, integrity, conscientiousness, neuroticism, and job performance. Measures of ability, conscientiousness, and neuroticism serve as predictors of task performance, and measures of integrity, conscientiousness, and neuroticism serve as predictors of contextual performance (Bacha, 2003; McHenry et al., 1990; Wise et al., 1990).

In order to predict aspects of task and contextual performance the composite developed will include scales from each of the predictors detailed in Figure 2.1. This should be an important consideration for any selection-focussed assessment or battery. The reason this is important is that task and contextual performance predictors predict different practical performance outcomes. For example, personality and ability measures might both predict 25% of future job performance for two employees, but the actual behaviour predicted may differ considerably. On the one hand, those with strengths in personality/integrity related areas might achieve their overall job performance by being more persistent, attentive, dependable, helpful, and so on. On the other hand, those with cognitive-based strengths are likely to achieve their overall job performance by solving problems relatively faster, effectively, accurately, and the like (Viswesvaran & Ones, 2002). Wood and Englert (2009) discuss the relevance of understanding these differences in practical performance within a selection context. To this end, Wood and Englert outline an intelligence compensation theory (ICT) describing the potentially compensatory nature of characteristics associated with conscientiousness. According to a number of researchers (e.g., Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Wood & Englert, 2009), less intelligent people can potentially compensate in the workplace by displaying more conscientious behaviour. Including task and contextual focused predictors helps provide comprehensive coverage of the job performance domain. It also reduces the likelihood that candidates assessed for potential recruitment are prematurely excluded due to a lack of information about other compensatory characteristics.

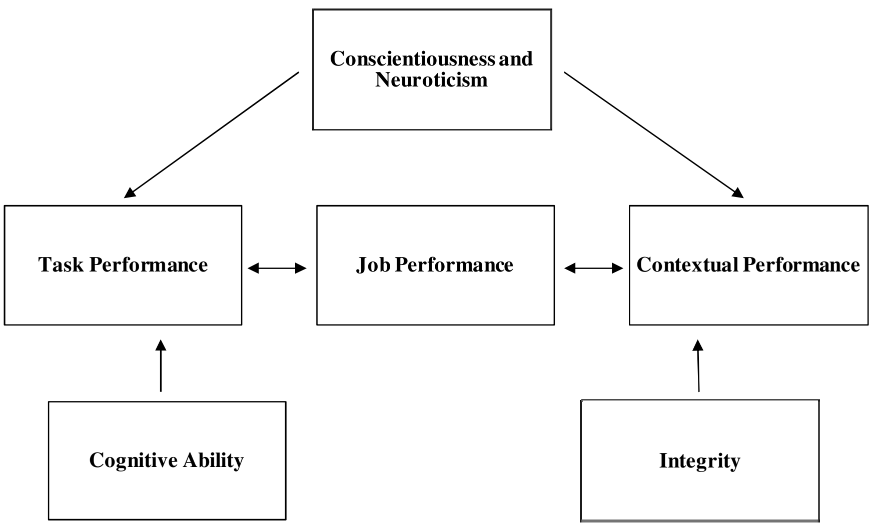


Figure 2.1. Literature Review Based Relationships between Performance Predictors and the Performance Domain

A recent focus of research into contextual performance has been within the sub-constructs of *counterproductive workplace behaviour* (CWB) and *organisational citizenship behaviour* (OCB). Organisational citizenship behaviours have been categorised in a variety of ways. The approach most commonly employed in contemporary research breaks OCB down into either individual or organisational targeted citizenship behaviours (OCBI and OCBO). This differentiates OCB behaviour according to its intended beneficiary. OCBI behaviours are those directed toward the benefit of other individuals. Examples of OCBI include altruism, peacekeeping, courtesy, and cheerleading behaviours aimed at helping other individuals. OCBO behaviours are those directed toward the benefit of the organisation. These include behaviours such as civic virtue, compliance, and organisational loyalty (Ilies, Scott, & Judge, 2006; Podsakoff, Whiting, Podsakoff, & Blume, 2009).

Counterproductive workplace behaviours (CWBs) are intentional behaviours that cause harm or have the potential to cause harm (Greenberg, 2003; Rotundo & Sackett, 2002). In other words, counterproductive behaviour is any intentional behaviour that an organisation deems to run counter to its legitimate interests. Researchers in the CWB domain have also differentiated between interpersonal and organisationally oriented CWBs. Like this differentiation within OCB research, these distinctions depend upon whether or not individuals or the organisation are the focus of these behaviours (Sackett, Berry, Wiemann, & Laczo, 2006). Individual CWBs encompass behaviours such as acting rudely or otherwise bullying colleagues (Sackett & DeVore, 2001). Organisation CWBs reflect behaviours such as deliberately producing poor quality work or theft (Bennett & Robinson, 2000).

The relevance of the organisational versus individual distinction lies in findings that OCBI, OCBO, and interpersonal and organisational CWB may be predicted by different personality traits. For example, OCBO may be more strongly predicted by conscientiousness (Ilies, Fulmer, Spitzmuller, & Johnson, 2009), while OCBI may be more closely aligned with neuroticism (Dalal et al., 2009). In similar findings organisational CWBs appear more strongly related to conscientiousness than interpersonal CWBs (Berry, Ones, & Sackett, 2007; Sackett et al., 2006), which may in turn be more strongly related to negative affectivity (Hershcovis et al., 2007).

The conceptual and empirical similarities between OCB and CWB domains have led some theorists to question whether these really reflect distinct constructs, or simply opposite ends of a single continuum (Collins & Griffin, 1998; Giacalone & Greenberg, 1997). This suggestion has received mixed support in some instances (e.g., Berry, Ones, & Sackett, 2007; Dalal, 2005; Dalal et al., 2009; Kelloway, Loughlin, Barling, & Nault, 2002; Sackett et al., 2006). Others have argued for their simultaneous or sequential co-occurrence (Sackett & Lievens, 2008; Spector & Fox, 2010a, 2010b). Fortunately for the goals of this dissertation, knowledge of the exact relationship between OCB and CWB is unnecessary for studying their shared antecedents and making predictions of more general contextual performance (Dalal, 2005; Levine, 2010; Sackett, 2002). Section 2.1.5 returns to questions concerning how well the proposed personality and ability predictors may account for unique variance within the job performance domain.

Section 2.1.5: Combining Predictors

The preceding subsections have suggested that intelligence, conscientiousness, neuroticism, and integrity are appropriate individual predictors of future job performance likely to collectively provide good coverage of the job performance domain. Incremental validity calculations provide evidence as to whether an assessment will *add* or *increase* the validity of predictions already made on the basis of other assessments employed (Sechrest, 1963). The focus of incremental validity within a psychometric context is on the value of adding new test data into a statistical equation intended to predict a criterion (Hunsley & Meyer, 2003). At the time of Ones et al.'s (1993) extensive meta-analytic review of integrity test validity a fundamental question remained unanswered. What incremental gain in validity results from adding integrity to intelligence tests in predicting job performance? Available studies suggested zero correlation between integrity and intelligence. On this basis Ones et al. calculated the addition of integrity testing to greatly increase the predictive validity of general mental ability alone. Subsequent reviews of integrity test validity have supported Ones et al.'s assumption that integrity and intelligence tests are not significantly related (Cullen & Sackett, 2004; Ones & Viswesvaran, 1998a; Sackett & Wanek, 1996).

Continued support for the hypothesis that integrity measures are essentially uncorrelated with intelligence is particularly useful in light of the emerging meta-analytic evidence for integrity tests' relationship with both measures of counter productivity and measures of overall job performance. This dual relationship has positive implications for the incremental validity of composites employing an integrity predictor in conjunction with intelligence (Wanek, 1999).

Estimates of integrity and intelligence incremental validity depends on job complexity. According to Ones et al. (1993) the effect of combining integrity and intelligence measures is an increase in predictive validity ranging from .22 (high complexity) to 1.04 (low complexity). Such differences are a consequence of integrity validity remaining constant over job complexity level while intelligence validity increases. As a result, the greatest gain in incremental validity as a function of combining integrity and intelligence measures appears for lower levels of job complexity. One of the limitations of this finding is that the integrity tests examined tended to be overt measures designed for lower complexity role selection. This may limit the accuracy of the finding for higher complexity roles.

A subsequent review by Schmidt and Hunter (1998) calculated a 27% increase in predictive validity when an integrity assessment is combined with cognitive ability. Schmidt and Hunter contend that the validity of a selection procedure is directly proportional to the procedure's utility. In other words, the validity of adding a predictor to a composite must be based upon the cost of that predictor relative to any increases in the performance of hired employees. The 27% increase in predictive validity they report for integrity including composites over cognitive ability alone could therefore be viewed as a 27% increase in utility.

Schmidt and Hunter (1998) have also calculated incremental validity for composites of intelligence and conscientiousness. Conscientiousness was reported as having a .31 correlation with Schmidt and Hunter's overall criterion of job performance. When added to the .51 correlation of intelligence an 18% increase in predictive validity was observed (.09 validity gain). The preceding investigations provide tentative support for the likely incremental validity of a composite including ability, personality, and integrity components. Yet further work will be required before greater certainty is achieved. Another consideration for a composite predictor of job performance is the issue of adverse impact. This is examined in section 2.1.6.

Section 2.1.6: Adverse impact

The adverse impact of intelligence tests upon minority groups is well documented. In the United States the use of cognitive ability tests in personnel selection has been shown to negatively impact upon African American and Hispanic job applicants. This is due to these groups having substantially lower mean ability scores than other applicants (Grubb et al., 2004; Sackett, Schmitt, Ellingson, & Kabin, 2001). Large scale meta-analytic investigations of ethnic differences in intelligence test means have confirmed that African Americans generally score about one standard deviation lower than majority group members on tests/scales of quantitative ability, verbal reasoning, and comprehension (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Roth et al. also found similar, though slightly smaller differences between majority group members and Hispanic job applicants (approx. .7-.8 standard deviations). A subsequent report released by the Scientific Affairs Committee (as cited in Guenole et al., 2003) estimated that those belonging to the minority groups of African Americans or Hispanics are the recipients of fewer opportunities for employment than their counterparts in certain other minorities or the White majority when selection decisions are based upon intelligence test scores.

It is a dilemma for many hiring organisations that intelligence tests are such effective predictors of future job performance, but have a discriminatory impact upon the hiring of some minorities. Organisations are well served employing intelligence tests for the maximisation of predictive validity. Yet in doing so they may be reducing the success of other objectives, such as overcoming historical social injustices, and creating both a pluralistic and ethnically diverse workplace and workforce (Sackett & Wilk, 1994; Sackett et al., 2001).

The literature search undertaken suggested that most research into the disparity among ethnic groups and workplace opportunities as a function of ability scores has occurred within the U.S.A. Yet potential selection disparities within the New Zealand workplace have been the focus of at least one investigation (Guenole et al., 2003). Guenole et al. examined verbal and numerical reasoning test scores for a participant group of 157 Māori and 82 European employees within a government organisation. Guenole et al. found that those identifying as Māori scored on average .55 standard deviations lower than their European counterparts on a measure of verbal reasoning. While finding no substantial mean difference between ethnicities on a test of general numeric reasoning, Māori scored a mean average standard deviation 1.79 lower than Europeans on a measure of numerical business analysis. While these findings failed to control for influences based upon education and experience, they do suggest that Māori may experience fewer opportunities within the workplace when selection decisions focus solely on ability outcomes.

Many organisations in New Zealand strive to achieve cultural diversity. Public Sector organisations have commitments to the State Sector Act (1988) and all organisations have obligations to honour the principle of partnership within the Treaty of Waitangi (1975). The State Sector Act 1988 requires each organisation covered by it to act as a "good employer." A "good employer" is defined as "an employer who operates a personnel policy containing provisions (…) requiring (…) recognition of the aims and aspirations and employment requirements, and the cultural differences, of cultural or minority groups" (s56 (2)). This is supported by provisions for the "impartial selection of suitably qualified persons for appointment." It is best to read these guidelines in conjunction with State Sector recruitment frameworks. They are not affirmative action requirements, but challenge State Sector recruiters to give ethnic minorities an equal chance in the recruitment process and ensure compliance with the State Sector Act (1988). The need to promote diversity while employing the best candidates increases the importance of developing strategies that will continue to allow for such maximisation via intelligence testing while minimising the adverse impact of such testing (Guenole et al., 2003). Yet this is no simple task. In fact Campion et al. (2001) consider the achievement of these often competing goals "the most perplexing problem facing the practice of personnel selection today" (p.158).

Guenole et al. (2003) have proffered a number of potential strategies by which organisations can continue to reap the benefits of testing ability in workplace selection while minimising the adverse impact of such tests upon Māori. Guenole et al. detail strategies such as the use of within-group score adjustments, score banding, and the use of non-cognitive selection alternatives. Each of these strategies go some way to achieving the goal of reducing adverse impact upon minorities and allowing for the maximisation of future job performance prediction. However, they also each have their drawbacks.

Within-group score adjustments fail to maximise predictive validity. They are not as effective as top-down selection from the entire group of applicants. Within-group score adjustments also invoke their own controversy as appointment decisions are at least in part based upon ethnicity rather than merit (Campion et al., 2001). The logic upon which score banding is based has also come in for criticism (Bobko & Roth, 2004; Campion et al., 2001). Score banding has also been criticised for failing to comply with the most fundamental premise of occupational testing – the optimisation of job performance prediction (Bobko & Roth, 2004; Campion, et al., 2001). This failure means that any reduction in adverse impact achieved by banding comes at the expense of the optimal use of an intelligence measure to predict performance, and thus the cost-effectiveness of such selection practices (Guenole et al., 2003). The strategy of including selection methods tapping into non-cognitive, job-related constructs, such as integrity tests is consistent with the modern emphasis of such tests in increasing predictive validity in selection decisions (Ilgen & Pulakos, 1999). Guenole et al. note that the inclusion of measures of non-cognitive aspects of job performance in selection processes can both increase validity and reduce adverse impact. However, the cost of failing to include cognitive measures, which have such well-established predictive validity for future job performance (Schmidt & Hunter, 1998), limits the utility of this strategy. Although it will not totally eliminate the adverse impact of intelligence tests, combining measures without significant ethnic bias that predict job performance should reduce such adverse impacts, whilst still optimising predictive validity (Sackett & Ellingson, 1997; De Corte et al., 2008).

A variety of researchers have investigated the impact of integrity testing on ethnic minorities. They have found overt integrity tests to have no adverse impact upon such groups (Sackett et al., 1989; Strand & Strand, 1986; Terris, 1983, 1985; Viswesvaran & Ones, 1997). Ones and Viswesvaran (1998a) also reported a variety of U.S.A based racial comparisons in integrity scores suggesting insignificant differences between African Americans and Whites, Hispanics and Whites, Asians and Whites, and American Indians and Whites. African Americans scored .04 standard deviations less than Whites. Hispanics scored .14 standard deviations higher than Whites. Asians scored .04 standard deviations higher than Whites. American Indians scored .08 standard deviations higher than Whites. As pointed out by Wanek (1999), these differences are in keeping with Cohen's (1988) interpretation that standard deviation units of .20 or less are small. Based upon the size of such differences the addition of integrity tests to selection processes involving ability measures could reduce the adverse impact of such measures substantially (Sackett & Ellingson, 1997).

Researchers have also examined more general personality traits for ethnic differences. Goldberg, Sweeney, Merenda, and Hughes (1998) explored the influence of ethnicity on Big Five personality within a participant group of 3629 individuals selected to be representative of the U.S.A. working population. Goldberg et al. found a trivial tendency for African Americans and Hispanic/Latino Americans to describe themselves as slightly less conscientious than Caucasian Americans (respectively $r = -.24$, $p < .01$; $r = -.23$, $p < .01$).

A number of other investigations have also examined the relationship between ethnicity and personality outcomes at the facet-level (e.g., Dion & Yee, 1987; McCrae, Yik, Trapnell, Bond, & Paulhus, 1998; Ones & Anderson, 2002). Although parallels with Big Five trait facets can be drawn, the facets examined in these investigations are not always representative of standard Big Five conceptions. For example, Asians score significantly higher than Europeans on the traits of social recognition, order, and harm-avoidance (Dion & Yee, 1987). The most relevant research undertaken concerning this dissertation's aims is Packman, Brown, Englert, Sisarich, and Bauer's (2005) investigation into differences in personality traits across ethnic groups within New Zealand. Packman et al. found significant effect size differences between Māori and New Zealand Europeans on the Big Five traits of extroversion ($d = -.24$, $p < .05$) and neuroticism ($d = .30$, $p < .05$). At the facet-level they reported significant effect size differences between Māori and New Zealand Europeans for intellectance ($d = -.43$, $p < .01$), enthusiastic ($d = -.29$, $p < .01$), and suspicious ($d = .54$, $p < .01$).

Information on ethnic differences in test scores is important for predicting disparate hiring ratios and other bias related outcomes. However, to facilitate incremental validity and calculate adverse impact reduction through the combination of predictors it is first necessary to know the relationships between those predictors. Relationships amongst ability, integrity, and personality are explored in subsequent sections and further clarified in the two studies undertaken within this dissertation. Section 2.2 examines the integrity construct, its measurement, and relationship with job performance.

Section 2.2: Integrity

"*As soon as the ability to engage in deception came into existence among our ancestors, there would have been selection pressure for detecting other people's efforts at deception*."

- Stewart-Williams (2004, p.217).

Kevin Murphy (1993) starts his book on honesty in the workplace by recounting the tale of the ancient Athenian philosopher Diogenes. According to legend, Diogenes wandered the Athenian streets carrying a lantern. This lantern was used to illuminate each stranger's face in a hopeful, yet fruitless search for an honest man. Murphy's point was that little had changed: scams, scandals, and the scoundrels responsible for them continue to abound. It is fair to say that little appears to have changed in the nearly two decades since Murphy's publication. Over the last few years the New Zealand media have brought to the public's attention numerous accounts of dishonesty in the workplace. Some of the more memorable have involved fictitious CVs, the misuse of public funds and political influence, and large-scale embezzlement. Yet based upon overseas research the biggest dishonesty costs within the New Zealand workplace are those unlikely to make the news. The most costly episodes of dishonesty are instead likely to involve low-level employee theft, the misuse of working hours, and employee absenteeism (Camara & Schneider, 1994; Vardi & Weitz, 2004).

As illustrated by Diogenes' search for an honest man, the need to detect deception is not a consequence of contemporary living. Although speculative, the role of dishonesty and its identification in human evolution is considered by many evolutionary psychologists to have played a crucial role in providing the impetus for the increased brain size in humans relative to other primates (Byrne & Whiten, 1988; Barrett, Dunbar, & Lycett, 2002).

Integrity tests for optimising employee selection through the prediction of dishonesty came into existence in the late 1940s (Ones & Viswesvaran, 1998a). Yet it wasn't until the introduction of the 1988 Employee Polygraph Protection Act in the United States of America that the use of integrity tests really took off (Cullen & Sackett, 2004). In 1989 O'Bannon, Goldinger, and Appleby reported that more than 5000 American employers administered more than 2.5 million integrity tests per year. Particularly high concentrations of integrity test use were reported in the retail sales, banking, and food service industries. Just one year later that figure had grown to 6,000 organisations using 5 million integrity tests annually (Camara & Schneider, 1994). A subsequent survey reported that around 14% of American companies were using integrity tests (Ones & Viswesvaran, 1998a). This is unsurprising considering the cost of employee theft alone for American businesses is estimated to be in excess of $200 billion per year (Penney, Spector, & Fox, 2003).

There are good reasons to suppose that figures for the use of integrity tests will now be even higher both internationally and within America. Perhaps first amongst these is the growing awareness of the cost and extent of counterproductive workplace behaviours. While integrity test use is still relatively novel within the New Zealand workplace, previous comments concerning the steady increase of counterproductive behaviour in the workplace internationally appear equally applicable within New Zealand (P. Englert, personal communication, January 22, 2010). Section 2.2.1 explores the construct of integrity and its assessment.

Section 2.2.1: The Integrity Construct

Integrity can be broadly conceptualised as an aspect of honesty manifest in one's adherence to a code of ethics or set of values (Branden, 1995). Integrity is commonly used to denote coherence and consistency concerning the honesty of one's words and actions; summed up by the idea that you will do what you say you will do (Kouzes & Posner, 1993; Shaw, 1997). Unfortunately an explanation of the construct of integrity measured by integrity tests is not so straightforward.

One of the many difficulties in the explication of the integrity construct is the variety of labels assigned to the constructs putatively measured by integrity tests. Camara and Schneider (1994) found publisher self-reports of the constructs measured by integrity tests to include the following: counterproductivity, honesty, job performance, attitudes, integrity, reliability, absenteeism/tardiness, admissions of dishonesty and drug abuse, credibility, dependability/conscientiousness, neuroticism, managerial/sales/clerical potential, predictiveness, probability of voluntary short-term turnover, service orientation, stress tolerance, substance abuse, substance abuse resistance, and supervisor ratings of performance. Camara and Schneider (1994) theorised that these constructs were subordinate to the higher-level construct of "trustworthiness."

Hogan and Hogan (1989) argue that the construct underlying integrity testing is really representative of mild tendency towards antisocial behaviour. According to this conceptualisation antisocial behaviour forms a syndrome comprising hostility toward authority, impulsiveness, social insensitivity, and feelings of alienation. All of these characteristics predispose people to defying rules, ignoring social expectations, and avoiding commitments to other people and organisations. These are normally distributed tendencies and those on the extreme end usually run afoul of public authority. According to Staw, Bell, and Clausen (1986) individuals with a mild tendency instead have careers marked by frequent job change, job dissatisfaction, and limited achievement (*cf.* Babiak & Hare, 2006). Collins and Schmidt (1993) choose to label the construct underlying the differences between non-offenders and those convicted of white-collar crime as "social-conscientiousness." Yet on the basis of the negative behaviours identified in integrity tests representing mid-range psychopathy, Hogan and Hogan prefer to label the construct underlying integrity tests "organisational delinquency." Such behaviourally focused conceptualisations receive further support from a finding that integrity test scores correlate strongly with assessments of a psychopathic personality, but not moral reasoning ability (Connelly, Lilienfeld, & Schmeelk, 2006).

Mehrabian (2000) also considers the construct measured by integrity tests to be a generalised predisposition concerning reliability, trustworthiness, honesty, dependability, and sociality. Yet Mehrabian resists providing any additional label. The nomenclature employed to describe the construct underlying integrity often differs according to the author proposing it. Nevertheless, there appears to be a great deal of similarity among the components comprising these constructs. Examining common features is therefore a good place to start in attempting to clarify the construct(s) measured by integrity tests. O'Bannon et al.'s (1989) review found integrity tests to generally include items relating to one or more of the following four areas:

- Direct admissions of illegal or questionable behaviour.
- Attitudes toward illegal or questionable behaviour.
- General thought patterns and personality traits thought to be related to dishonesty. For instance, the tendency to dwell upon illegal activities.
- Responses to hypothetical situations involving honest or dishonest behaviour.

Murphy (1993) has detailed seven dimensions that are often reported in factor analyses of integrity tests:

- The first concerns perceived incidence of dishonesty. This is based upon the idea that less honest individuals are likely to estimate a higher incidence of dishonest behaviour.
- The second concerns leniency toward dishonest behaviour. This is based upon the idea that less honest individuals are more likely to forgive or excuse dishonest behaviour in the workplace.
- The third concerns theft rationalisations. This is based upon the idea that less honest individuals are likely to come up with more excuses or reasons for theft.
- The fourth dimension concerns theft temptation or rumination. This is based upon the idea that less honest individuals are more likely to think about theft.
- The fifth concerns norms regarding dishonest behaviour. This is based upon the idea that less honest individuals are likely to view dishonest behaviour as acceptable.

- The sixth concerns impulse control. This is based upon the idea that less honest individuals are likely to act on their impulses.
- The seventh concerns punitiveness towards oneself or others. This is based upon the idea that less honest individuals are more likely to have more punitive attitudes.

A meta-analysis (180 studies, 665 validity coefficients) of different integrity tests undertaken by Ones, Viswesvaran, and Schmidt (1993) found an underlying similarity among such tests. This prompted them to question whether integrity tests didn't all primarily measure a single superordinate construct. Ones et al. theorised that the most likely candidate was the broadly defined general construct of conscientiousness. Conscientiousness is one of the five hypothesised personality constructs within the Big Five model of personality (Barrick & Mount, 1991; Digman, 1990; Goldberg, 1990). Conscientiousness generally reflects such characteristics as carefulness, self-control, dependability, and responsibility. The Big Five model and its traits are thoroughly explicated in the personality section of this chapter. According to Ones et al. the major difference between the construct of conscientiousness in personality and integrity literature is the end of construct's continuum measured. In other words, personality assessments examine the positive components of conscientiousness (e.g., responsibility, carefulness, and rule following), while integrity assessments measure the negative behaviours associated with a lack of conscientiousness (e.g., irresponsibility, carelessness, and violation of rules). Subsequent reviews have further clarified this suspected relationship between conscientiousness and integrity tests with varying degrees of success (Sackett & Wanek, 1996). Despite such variation, the hypothesis that conscientiousness scores will be correlated with those of integrity tests is clearly appropriate. What is known of the relationship between integrity and conscientiousness is further detailed in section 2.3.3.

A later investigation by Wanek, Sackett, and Ones (2003) identified 23 thematic composites on the basis of a judgmental sort of 798 items from seven paper-and-pencil integrity tests. A principal components analysis of these 23 composites identified four principal components underlying the integrity construct. Wanek et al. labelled their principal components and described the focus of each as follows:

Principal component 1 – *Antisocial behaviour*: this component focuses upon theft, breaking rules, and wrongdoing.

Principal component 2 – *Broad socialisation*: this component had strong loadings upon achievement/success orientation, locus of control, neuroticism, extroversion, and positive home life/upbringing.

Principal component 3 – *Positive outlook*: this component is dominated by the composite of safety/accident proneness, and involves perceptions of the prevalence of dishonesty and attitudes towards supervision.

Principal component 4 – *Orderliness/diligence*: this component is labelled to reflect its strong loadings upon the composites of orderliness and diligence.

Wanek et al. (2003) identified these principal components as those underlying the construct of integrity. This suggests that while various integrity tests may be quite different in terms of their surface content, they are assessing the same underlying constructs. A limitation of this finding is that principal component analysis was used to identify how many discrete components exist, rather than how many components should be included in a subsequent factor analysis. Such a factor analysis would have provided greater confidence in the discrete existence of the four components identified (Kline, 1994). However, Wanek et al.'s finding was closely paralleled in Green's (2003) research into the underlying dimensions of integrity that did use factor analyses. As a result of confirmatory and discriminant analyses Green labelled the four primary factors underpinning the second-order factor of integrity as: concern for others, consciousness, honesty, and emotional control.

A serious limitation of the majority of integrity related research previously discussed and explored to better understand the integrity construct is their reliance on integrity measures rather than the behavioural outcomes they are intended to predict. On this basis a potentially more useful way to understand the similarities among integrity tests and their underlying constructs is via the broad category of counterproductive workplace behaviour (CWB). As previously discussed, counterproductive workplace behaviour is in many respects the antithesis of organisational citizenship behaviours. Counterproductive behaviour encompasses the diverse areas of intrapersonal misbehaviour, interpersonal misbehaviour, production misbehaviour, property misbehaviour, and political misbehaviour (Vardi & Weitz, 2004). Counterproductive workplace behaviour appears to cover a sufficiently broad range of behaviours to capture all previous conceptualisations of what the integrity construct is attempting to measure. Further support for use of the CWB framework comes from findings suggesting that as the likelihood an individual will engage in any one type of CWB increases, so does the likelihood that they will engage in other forms of CWB (Gruys & Sackett, 2003; Viswesvaran, 2002).

While disagreement as to the most useful taxonomy for the content of integrity tests has not been uncommon, there is much greater consensus for their broad categorisation as belonging to one of two categories of measure. Sackett, Burris, and Callahan (1989) have labelled these categories "overt" and "personality-oriented" tests. Murphy (1993) has provided the alternative labels of "clear-purpose" and "veiled-purpose" tests. Sackett and Wanek (1996) found that overt integrity tests are generally comprised of two sections. The first section typically measures an applicant's attitudes towards theft. This section contains questions relating to beliefs about the extent and commonality of theft, the supposed ease of theft, punitiveness toward those who steal, support for frequent rationalisations for theft, ruminations about theft, and judgments concerning one's own honesty. The second section of overt integrity tests is comprised of requests for admissions of theft and other deviant behaviour. In this second section applicants are invited to describe the amount and frequency of theft and other illegal and/or CWBs they have engaged in. According to Sackett and Wanek (1996) these two sections of integrity tests are also occasionally packaged with other scales. Such additional scales are designed to measure factors such as tendencies toward violence and drug abuse. Commonly used overt tests of integrity include the London House PSI, the Reid Report, and the Stanton Survey. The first study of this dissertation employs the Stanton Survey (see section 3.1.3).

Personality-oriented tests are closely associated with non-clinical measures of personality, such as Cattell's 16PF personality inventory. These are considerably broader in focus than overt tests. They are also not explicitly directed towards CWBs. Personality-oriented tests are comprised of items relating to conscientiousness, dependability, social conformity, thrill seeking, hostility, and problems with authority. Commonly used personality-oriented instruments include the Reliability Scale of the Hogan Personality Series, the PDI Employment Inventory, and the Personnel Reaction Blank (Sackett & Wanek, 1996). Personality-oriented tests of this type are not uncommon, but a more thorough examination of their composition is beyond the scope of this investigation. This dissertation restricts its focus to overt tests of integrity. Concentrating on overt integrity tests avoids the conceptual confusion that might otherwise result from comparing personality-oriented tests of integrity with other non-integrity based tests of personality. Overt tests are also considered better predictors of CWBs than personality-based measures (Goodstein & Lanyon, 1999; Rafilson & Frost, 1989; Sackett & Wanek, 1996). Overt integrity tests were also found to have superior psychometric properties concerning reliability and more general validity (Ones & Viswesvaran, 1998b). For further information on personality-oriented tests of integrity see Ones et al. (1993b) or Hogan and Brinkmeyer (1997).

Section 2.2.2: Integrity Test Prediction of Job Performance

Evidence indicates that applicants who score poorly on integrity tests make worse employees. It is not necessarily that these individuals become involved in theft or other kinds of serious wrongdoing, but that they are relatively more likely to be irresponsible, undependable, late, absent, and so on (Ones et al., 1991). This suggests integrity testing can assist in improving personnel selection decisions by screening out potentially counterproductive employees.

The relevance of integrity tests to comparatively narrow criteria such as theft or other types of property deviance appears straightforward. Yet the broader relationship between integrity tests and global measures of job performance is not quite so obvious (Murphy, 1993). One explanation is that what integrity tests actually predict amounts to poor job performance. In particular, counterproductive behaviours such as irresponsibility and reliability are themselves components of poor job performance. This is consistent with findings that counterproductive behaviours such as lateness, absenteeism, and turnover are all significantly related to integrity test scores (Johns, 2002; Simonini, 1998). This is also consistent with the previously discussed conceptualisation of the integrity domain concerning CWBs. As integrity tests are designed to predict these and other counterproductive behaviours they ought naturally to indicate likely job performance (Murphy, 1993). A number of meta-analytic investigations have found positive correlations between integrity tests and supervisory ratings (e.g., Berry, Sackett, & Wiemann, 2007; Ones et al., 1993, Sackett & Wanek, 1996). It is hypothesised that this relationship is due to work supervisors considering counterproductive behaviour when forming an overall evaluation of an individual's performance. This would explain integrity's relationship with general job performance even though integrity test scores are focused on counterproductive work behaviours. This is particularly likely if scores moving away from the low integrity end of the continuum are related to an increase in organisational citizenship behaviours.

Murphy and Lee (1994a) have suggested an alternative explanation for the correlation between integrity test scores and job performance. Murphy and Lee observed that the majority of integrity tests appear to either directly or indirectly assess conscientiousness. This observation was earlier touched upon within the context of construct clarification (Ones et al., 1993). Conscientiousness is strongly positively correlated with job performance across an unusually wide variety of jobs (Barrick & Mount, 1991). According to Murphy (1993) it is likely that the positive relationships between conscientiousness and integrity test scores are partly responsible for the correlation between integrity test scores and job performance. Figure 2.2 illustrates these potential relationships and interactions.

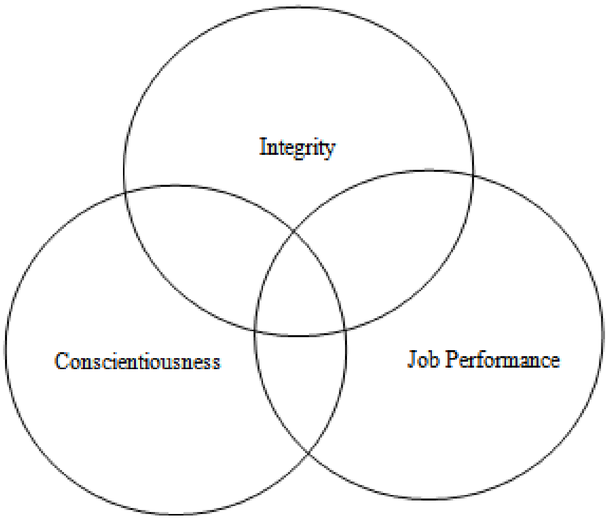


Figure 2.2. Integrity, Conscientiousness, and Job Performance Relationships (adapted from Murphy, 1993)

According to Murphy (1993) the best way in which to understand this hypothesis is by thinking about individuals exhibiting low levels of conscientiousness. These people avoid extra work (often even their ordinary work if they are able to get away with it), are undependable and sloppy, and in need of constant close supervision. This is not exactly the profile of a model employee.

Murphy and Lee (1994a) collected data from two participant groups to test the hypothesis that conscientiousness either partially or fully explains the correlation between job performance and integrity. Using existing meta-analyses (Barrick & Mount, 1991; Ones, Viswesvaran, & Schmidt, 1992), Murphy and Lee were in a position to estimate the relationships between job performance and integrity, and between job performance and conscientiousness. The goal of their investigation was to estimate the extent of the correlation between conscientiousness and integrity. The outcome of this investigation indicated that both overt and personality-oriented integrity test scores are positively correlated with measures of conscientiousness. It also indicated that such relationships are not sufficiently strong to account for the relationship between job performance and integrity test scores. Murphy and Lee's investigation appears to indicate that the personality trait conscientiousness provides only a partial explanation for the relationship between job performance and integrity. There are factors other than conscientiousness requiring consideration in order to understand the correlation between integrity and job performance. Murphy (1993) did not feel confident at the time of this evaluation that those factors could be easily determined. It is hoped that this dissertation's examination of integrity test score correlations with personality trait facets, and fluid and crystallised intelligence, will contribute to our understanding of what factors other than conscientiousness might explain integrity test outcomes.

Sackett and Wanek (1996) reached a similar conclusion for the relationship between integrity and job performance. Like Murphy (1993), Sackett and Wanek (1996) hypothesised that integrity tests are not only linked to counterproductive work behaviours. They also believed that integrity tests reflect underlying personality traits through which integrity is linked to positive job performance. Conscientiousness was again the primary candidate.

According to this hypothesis, people scoring high on integrity tests not only avoid CWBs, but also engage in persistent, careful, and productive behaviour. Sackett and Wanek (1996) presented ten correlations derived from objective production records in support of this personality trait explanation for the relationship between integrity and job performance. They contend that if the relationship between integrity test scores and job performance is a function of supervisor's consideration of CWBs in overall performance ratings, then the expected relationship between integrity test scores and objective production records would be zero. They instead found the mean correlation with production records virtually identical to those involving supervisory ratings. This is consistent with their personality trait linked explanation. That is, this finding appears to support the contention that integrity tests are linked to workplace behaviours reflecting factors other than counterproductive behaviour. The relationship among personality traits such as conscientiousness and integrity will be examined in detail in the following personality section.

Regardless of why integrity relates to performance, many studies have supported the relationship between job performance and integrity test scores (e.g., Berman, 1993; Bernardin & Cooke, 1993; Borofsky, 1992; Borofsky & Smith, 1993; Collins & Schmidt, 1993; Jones Joy, Werner, & Orban, 1991; Jones, Slora & Boye, 1990; Kobs & Arvey, 1993; Kroeck & Fraser, 1991; Mikulay & Goffin, 1998; Ones & Viswesvaran, 1998b). Large-scale meta-analytic investigations have also supported integrity test validity (Ones et al., 1993; Sackett & Wanek, 1996). Ones et al. (1993)

concluded that integrity test validity is positive and in useful ranges for predicting both CWBs and overall job-performance criteria. Further support comes from a review by Ones, Viswesvaran, and Schmidt (1995) suggesting an integrity test operational validity of .41 in the prediction of overall job performance. A subsequent review replicated the finding that .41 was the average validity of integrity tests for supervisory ratings of job performance across settings (Ones & Viswesvaran, 1998b, 1998c).

An important limitation of all of these findings concerns the range restrictions associated with the roles examined. The vast majority of these investigations focused on the low complexity roles for which standard integrity tests are designed. This seriously limits the potential generalisability of these results.

With this limitation in mind, the preceding findings suggest that integrity tests have substantially higher predictive validity than many of the more accepted methods of personnel selection, including: biodata, assessment centre ratings, reference checks, and personality inventories (Schmidt & Hunter, 1998; Schmidt, Ones, & Hunter, 1992). As well as predicting job performance, integrity tests have been found to significantly predict training success, accidents in the workplace, property damage, withdrawal behaviour, violence, and drug and alcohol abuse (Ones & Viswesvaran, 1998b; Schmidt, Viswesvaran, & Ones, 1997). Although the preceding citations generally fail to take into account potentially shared variance for performance prediction amongst such scales, support for the validity of integrity tests in predicting work relevant behaviour still appears reasonable. Section 2.3 explores more general personality assessment.

Section 2.3: Personality

"Tradition has it that fat men are jolly and generous, that lean men are dour, that short men are aggressive, and that strong men are silent and confident".

- Sheldon (1942, p.1).

Our understanding of personality traits has progressed since Sheldon's constitutionally-based differences in personality (Sheldon & Stevens, 1942). Yet consensus as to its conceptualisation remains elusive. According to the Penguin dictionary of Psychology (Reber & Reber, 2001) the most useful way to conceptualise personality is in accordance with its role in different personality theories.

Ewen (2003) describes five major perspectives in personality theory: The psychodynamic, the humanistic, the behaviourist, the cognitive, and the trait. Psychodynamic theories emphasise the unconscious. They contend that many of personality's most important components are beyond our awareness and can only be made conscious with the greatest of difficultly, if at all (Freud, 1917; Horney, 1945; Jung, 1933). Humanistic theories emphasise our innate potential for healthy growth and development. According to this perspective, psychopathology is a consequence of our healthy potentials being blocked by harmful environmental forces, such as pathogenic parenting (Maslow, 1970; Rogers, 1961). Behaviourism aims at making psychology more scientific by limiting its focus to the observable (Skinner, 1965). In doing so behaviourism eliminates from consideration virtually all of what personality theorists usually consider important: inner causes of behaviour, needs, drives, thoughts, wishes, emotions, expectations, etc. This makes behaviourism more of an alternative to personality theory than an approach. Cognitive theorists emphasise the role of thinking in personality. Behaviour is determined by how we interpret, evaluate, and predict events, rather than by reality or any innate instincts (Bandura, 1977; Kelly, 1963). A *motivated cognition approach* to personality suggests that personality becomes manifest through motivated biases and preferences in the ways that people see the world and cope within it (Higgins & Scholer, 2008). The focus of this dissertation is the trait approach to personality.

Trait theorists emphasise a surface-oriented approach to personality based on empirical research rather than clinical observation. They describe the conscious and concrete aspects of personality in straightforward terms (e.g., "ambitiousness," "disciplined"). Unconscious and abstract explanations of human behaviour are de-emphasised by trait theorists (Allport, 1968; Cattell, 1946). Trait theorists conceive personality traits as those inner structures directing an individual's behaviour in a relatively typical and stable manner (e.g., Allport, 1968). Yet the focus of such a definition upon "inner structures" is potentially problematic (Cervone, 2005; Read, Monroe, Brownstein, Yang, Chopra, & Miller, 2010). A less contentious definition is that a personality trait reflects the nomenclature for a *descriptive* classification of behavioural/cognitive consistencies (Ewen, 2003).

The trait approach is consistent with the applied selection focus of the personality variables used within this investigation. It provides the paradigm of greatest currency within both applied and academic I.W.O. psychology (Roberts & Hogan, 2001). A further discussion of the trait paradigm, followed by a brief discussion of the theoretical background underpinning the operational definition of personality used in this research is contained in the next subsection. This discussion helps clarify the parameters of the term *personality* employed in this investigation. This is important in later comparisons of personality with the constructs of intelligence and integrity. This is followed by brief discussions of the relationship between personality and job performance, and a review of the literature concerning the relationship between personality traits and integrity. The former provides the rationale for the inclusion of a personality measure within this investigation. The latter facilitates an understanding of the issues surrounding construct similarities and differences. This is important to frame expectations concerning relationships between integrity and personality.

Section 2.3.1: Personality Traits

The origins of the trait model of personality are traceable to Gordon Allport (1937). Allport wanted to counter what he regarded as Freud's excessive focus on hidden meanings and motives. Allport saw personality as the organising force within an individual that is responsible for determining characteristic patterns of behaviour. Allport theorised that such patterns of behaviour are manifest in traits such as *ambitiousness* and *friendliness*. Allport believed in the value of identifying these traits. Yet he thought each individual's personality too unique to be accurately described by a general reference to such traits. Allport believed that people are motivated both by the desire to reduce some drives (e.g., hunger and thirst) and increase others (e.g., sex), but that the unconscious and defence mechanisms were only relevant in unhealthy personalities. Like other personality theorists Allport believed it was possible to seek deep-seated explanations for personality. Unlike other theorists Allport believed that personality description is the correct focus for psychology and deeper probing of personality is largely unnecessary (Allport, 1968). In order to determine the parameters of this idiographic theory of trait personality Allport and Odbert (1936) undertook an examination of the terms used to describe personality within the English language. This investigation uncovered 4,500 trait adjectives from which to describe an individual's personality.

The next real progress within the trait paradigm of personality theory had its genesis in the work of Raymond Cattell. Cattell strongly disagreed with Allport's (1937) idiographic approach to personality and instead proposed a nomothetic approach. Cattell's approach considered differences in personality sufficiently uniform as to allow the identification of broadly applicable descriptive traits. To this end Cattell (1943a) reduced Allport and Odbert's (1936) list of English trait adjectives from 4,500 to 171 through the elimination of perceived synonyms. Cattell then obtained observer ratings for individuals on these adjectives. Based on these ratings Cattell was able to distinguish 36 clusters of correlations, or surface traits. Cattell defined surface traits as visible personality characteristic manifestations of the combination of two or more source traits. Cattell considered source traits the basic elements of personality and solely discernable through factor analysis. Through the use of factor analysis Cattell was able to identify sixteen source personality traits (Cattell, 1946). A derived assessment of these traits is employed in the following chapter's exploration of predictor intercorrelations (see section 3.1.3.2).

One benefit of employing Cattell's theory of broadly descriptive traits is its consistency with the evolutionary approach to understanding human personality. According to Tooby and Cosmides (1990) the evolutionary constraints of selection pressures make personality types unavoidably unitary rather than individually unique. These selection pressures are shaped by the inherently social nature of reproduction and the constraints of genetic recombination. Such adaptive and genetic parameters preclude the development of the totally unique and distinct personality types proposed by Allport. Buss (1991) discusses a variety of publications consistent with Cattell's trait approach to personality. These articles have attempted to identify the particulars of this unitary human nature from an evolutionary perspective. In one example Wiggins (1990) argues that the motivational modes of agency (striving for power and mastery that distinguish one individual from another) and communion (striving for intimacy, union, and solidarity with others) provide the necessary conceptual framework within which to understand interpersonal behaviour via trait differences. In another example it is argued that the basic human motivators responsible for personality trait development and differences are status and popularity (Buss, 1991). In yet another it is proposed that the *anxiety* trait is a species-typical adaptation that provides motivation for preventing social exclusion (Baumeister & Tice, 1990).

The trait approach's consistency with the parameters of evolutionary theory is crucial to the legitimacy of its attempt to understand psychological mechanisms and manifestations (Dennett, 1995). Yet the parameters of trait theorists' conceptualisation of the personality construct have not been as consistent. I.W.O. psychologists originally conceptualised the construct of personality as an all-inclusive reference to any traits or characteristics able to account for an individual's behaviour in a given situation (Ryckman, 1993). The pioneers of personality research defined traits as relatively broad and permanent behaviour tendencies. The most important distinction among traits for such pioneers was between ability and non-ability traits (Ewen, 1988). Ability traits are those which are directly associated with one's capacity to perform specific functions and are generally conceptualised in terms of intelligence. Ability traits are value-laden in that their relatively greater possession is considered inherently beneficial (*cf*. Reader, 2004). Non-ability traits are predispositions to behave in certain ways that are not directly associated with one's capacity for performing some particular function. Unlike ability traits they are not inherently value-laden. This value neutrality is exemplified by the fact that diametrically opposed non-ability traits can be considered good or bad depending upon situational considerations. This ability/non-ability trait distinction contrasts sharply with contemporary conceptualisations of personality traits. Almost without exception contemporary conceptualisations consider "ability" to be something other than a personality trait (Ewen, 2003).

Like biological evolution, the evolution of the personality trait construct into something excluding ability appears a consequence of convenience and efficiency rather than design. This change in conceptualisation can be traced be to "whole person" theorists' recognition that dividing the construct of personality into its component parts would advance the cause of personality research (e.g., Cattell, 1957). Such compartmentalisation enabled investigators to select manageable components of personality for study (Ewen, 1988; Highhouse, 2002; Mischel, 1981; Saklofske & Zeidner, 1995). Yet the *artificial* nature of the treatment of intelligence and non-intelligence components of personality as independent constructs appears to have been largely forgotten by subsequent generations of researchers. As a result the term personality has come to be fairly consistently employed by researchers and practitioners as a construct describing a compendium of traits or characteristic ways of behaving, feeling, thinking, and reacting distinct from intelligence (e.g., Barratt, 1995; Buss & Plomin, 1975; Eysenck, 1970; Eysenck & Eysenck, 1985; Mischel, 1981; Necka, 2003). For the sake of convenience, it is this ubiquitous non-intelligence conceptualisation of personality that is used within this dissertation. This is not intended to endorse the artificial separation of ability and non-ability in conceptualisations of personality. The dissertation also often refers to integrity as a construct distinct from that of general personality. The rationale for this decision is primarily based upon its lack of inclusion in predominant models of personality. It is also based upon the similarity of integrity to intelligence test outcomes in respect of their value-laden nature. The next subsection examines the Big Five model of personality, which is the most commonly employed contemporary model of personality.

Section 2.3.1.1: The Big Five Personality Model

The emergence of a widely accepted personality model has provided applied psychologists with an unrivalled opportunity to organise personality trait constructs and to link these constructs to performance criteria (Kanfer et al., 1995). The introduction of the Big Five personality dimensions is often associated with McCrae and Costa's (1985) addition of the conscientiousness and agreeableness dimensions to the NEO Personality Inventory (NEO-PI). The NEO-PI previously comprised: neuroticism, extroversion, and openness-to-experience. Costa and McCrae's Big Five paradigm is consistent with a rapidly increasing body of research supporting the utility of five major personality traits (Costa, Herbst, & McCrae, 2002; Digman, 1990; Hough & Schneider, 1996; McCrae & Costa, 1985, 1987, 1997a; McCrae & John, 1992; Ones, Viswesvaran, & Dilchert, 2005b, 2005c). The replicability of these factors throughout adulthood has been supported by both cross-sectional and longitudinal studies (e.g., Costa & McCrae, 1988; Costa et al., 1986; McCrae et al., 2000). The stability of these traits has also been demonstrated across both cultures and genders (Costa, Terracciano, & McCrae, 2001; McCrae et al., 1999; McCrae, Costa, & Del Pilar, 1998; McCrae, Costa, & Terracciano, 2002; Ones & Anderson, 2002; *cf.* De Raad et al., 2010). Yet agreement concerning the Big Five's utility has not equated with consensus on *which* five factors it should emphasise (Matthews & Deary, 1998). Nor is it equated with agreement as to *how many* or *which* facets underpin each of the Big Five (Block, 2001). Although consensus is lacking, reviews by Digman (1990), Wiggins and Pincus (1992), and Hogan and Ones (1997) all include the following dimensions or their variants: extroversion/introversion (also called surgency), neuroticism (emotional stability), agreeableness (likeability), conscientiousness (conformity, self-control, dependability), and openness to experience (intellect, culture).

The Big Five factor model employed in the Chapter 4 investigation is comprised of the bi-polar traits detailed in Table 2.1. A high/positive score on one end of the continuum signifies a low/negative score on the other end of the continuum. To distinguish the level of trait focused upon this dissertation will refer to the primary order traits comprising each Big Five scale as facets. The facets of which these Big Five are comprised are detailed in Table 2.2.

Table 2.1

15FQ+ Big Five Model of Personality Bi-Polar Dimensions

Traits	Trait Descriptors	Opposite Traits	Opposite Descriptors
Extroversion	Oriented to the external world of people, events, and activities. Prefers social contact and stimulation.	Introversion	Oriented to inner world of thoughts, perceptions, and experiences. Not needing much social contact.
Stability	Calm, resilient, and well adjusted. Equipped for emotionally demanding situations. Unlikely to ruminate.	Neuroticism	Prone to mood swings, touchy, sensitive, and vulnerable. Difficulty coping.
Pragmatism	More influenced by tangible evidence and hard facts than subjective experience. Potentially insensitive to subtleties and people.	Openness	More influenced by new ideas, feelings and sensations than tangible evidence or hard facts. Open to possibilities and subjective experience/opinions.
Independence	Independent minded. Actively self-determined in own actions and thoughts. Often strong-willed, intractable, and confrontational.	Agreeableness	Tolerant, agreeable, and obliging. Not disagreeable, stubborn, nor opinionated, but happy to compromise.
Expedience	Less influenced by social norms or internalised parental expectations. Low levels of self-control and restraint. Flexible and able to view things from a broader perspective.	Conscientiousness	High degree of self-control and restraint influenced by internalised parental expectations and social norms. Detail orientated, rule-bound, and systematic and orderly.

Note. Derived from information contained within Psychometrics Ltd (2002). Where necessary, 15FQ+ trait labels are altered to conform to standard usage (e.g., "high self-control" is labelled "conscientiousness").

Section 2.3.2: Personality Trait Prediction of Job Performance

There is an extensive body of literature on the relationship between the Big Five and job performance measures (Hurtz & Donovan, 2000; Kroeck & Brown, 2004; Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett, Jackson, & Rothstein, 1991). Support for the validity of Big Five factors in the prediction of job performance comes from more than two dozen meta-analyses, spanning decades of research, and incorporating thousands of individual studies (Goodstein & Lanyon, 1999; Ones et al., 2005b, 2005c). One of the most widely cited of meta-analytic investigations of the relationship between the Big Five personality dimensions and job performance is Barrick and Mount (1991). Barrick and Mount examined 117 studies that reported correlations between personality tests and measures of job performance, training proficiency, and such characteristics as turnover, tenure, absenteeism, and salary level in professional, managerial, sales, police, and skilled/semiskilled jobs. Barrick and Mount's analyses indicated that conscientiousness is the only personality trait that is consistently correlated with job performance measures. Moreover, Barrick and Mount demonstrated that the validity of conscientiousness as a predictor of performance is essentially constant across different types of jobs and performance measures. Conscientiousness has repeatedly emerged in subsequent reviews as the most robust of Big Five predictors of job performance (Kroeck & Brown, 2004; Warr, Bartram, & Martin, 2005). Schmidt and Hunter's (1998) meta-analysis of the validity and utility of selection measures found conscientiousness the best Big Five trait for enhancing ability lead incremental validity. In fact, subsequent research has been so bold as to claim that conscientiousness is the motivational variable that I.W.O. psychologists have long sought (Barrick, Mount, & Judge, 2001).

Hurtz and Donovan (2000) explored how conscientiousness differed in its prediction of task and contextual performance. They did this by partitioning job performance into task performance and the contextual components of job dedication and interpersonal facilitation. True operational validity scores for conscientiousness on task performance were found to be .15, job dedication .18, and interpersonal facilitation .16. These are very small correlations, but do suggest that conscientiousness consistently predicts both contextual and task performance nearly equally well. A variety of other investigations have supported the role of conscientiousness in both task and contextual performance.

Judge et al.'s (2008) meta-analytic review found correlations between conscientiousness and job performance .28, job satisfication .26, leadership .28, motivation .28, and accidents -.31. At least three major meta-analytic investigations have also found conscientiousness to be the second most important task performance predictor (Barrick & Mount, 1991; Ones et al., 1993; Schmidt & Hunter, 1998). The ability of conscientiousness to predict behaviours related to contextual performance and organisational citizenship behaviour is also well supported. For instance, research by Hogan, Rybicki, Motowidlo, and Borman (1998) demonstrated a positive relationship between conscientiousness and contextual performance indicators such as work dedication and interpersonal facilitation. Other researchers have demonstrated conscientiousness' ability to predict organisational citizenship behaviours directed toward one's coworkers and organisation (Ladd & Henry, 2000; McNeely & Meglino, 1994; Organ, 1994; Organ & Ryan, 1995).

The Big Five trait of conscientiousness has also been shown to predict job performance across different stages of employment. Its facets have also been shown to predict different stages to different extents. Stewart (1999) found the facet of need for achievement to be most strongly correlated with job performance in the maintenance stage (veteran employees). He found the facet of order more strongly correlated with performance in the transitional stage (newly hired employees). The consistency of these findings and the ubiquity of conscientiousness assessment use make this trait an important predictor for inclusion in subsequent analyses.

Neuroticism has also been shown to be a potentially important Big Five predictor of job performance (Barrick, Mount, & Judge, 2001; Kroeck & Brown, 2004; Ones, Dilchert, Viswesvaran, & Judge, 2007). This may in part be due to neuroticism's apparent ability to account for as much as 10-25% of variance in job satisfaction (Connolly & Viswesvaran, 2000). Hurtz and Donovan's (2000) investigation found true operational validity scores for neuroticism on task performance to be -.13, job dedication -.13, and interpersonal facilitation -.16. Although the preceding correlations are very small, Hurtz and Donovan's results suggest neuroticism also consistently predicts both contextual and task performance nearly equally well. Judge et al.'s (2008) meta-analytic investigation found neuroticism to be correlated to job satisfaction -.29, leadership -.24, accidents .30, and motivation -.29. Other reviews examining the relationship between neuroticism and job performance have also provided support for the claim that neuroticism is a valid predictor of job performance (Barrick et al., 2001; Moscoso & Salgado, 2004; Tett et al., 1991; Salgado, 1997). These reviews place neuroticism in the number two spot behind conscientiousness for Big Five factors in the prediction of job performance. This was a position previously assigned to extroversion (Kroeck & Brown, 2004). These findings make neuroticism an important predictor for inclusion in subsequent analyses.

Extroversion has been shown to be a valid predictor of performance for sales personnel (e.g., Vinchur, Schippman, Switzer, & Roth, 1998) and managers (e.g., Furnham, Crump, & Whelan, 1997), though the correlation coefficients are typically below .20. Although more recent research has failed to replicate such findings, Mount and Barrick's (1995) review found extroversion validities to range between .13 and .51. Recent reviews by Salgado (1997) and Hurtz and Donovan (2000) yielded coefficients of .12 and .09 respectively. Due to such discrepancies among the magnitudes assigned to its relationship with performance the predictive validity of extroversion remains unclear.

Amongst Big Five factors agreeableness and openness to experience have typically been found to be weak general predictors of job performance. For example, Barrick and Mount's (1991) meta-analysis yielded a coefficient of only .07 for the relationship between job performance and agreeableness. With the exception of Tett et al.'s (1991) meta-analysis, openness to experience has been consistently found to have the lowest validities of Big Five factors in the prediction of job performance. Moreover, research examining situational variation in Big Five traits for predicting sales performance has found the validity of agreeableness highly situationally specific (Warr et al., 2005). Mount, Barrick, and Stewart (1998) found an average association of .20 between agreeableness and rated overall job performance in close team-working situations. Yet Crant (1995) reported a correlation of -.11 between agreeableness and objective sales among commission-earning real estate agents. These contrary findings suggest that the relationship between job performance and agreeableness is more role specific/dependent that the more generalisable results reported for conscientiousness and neuroticism. Although these estimates may in themselves be underestimates based on combing differing results within a meta-analytic context.

One of the factors likely to reduce the accuracy of predictive relationships between those personality traits described and job performance is the failure to take into account complex trait interactions (Wood & Englert, 2009). For example, results suggest that having strong tendencies towards both social interaction (extroversion) and fear of rejection/embarrassment (neuroticism) can negatively impact upon motivation and performance (Robinson, Wilkowski, & Meier, 2008). Other research suggests highly conscientious employees who lack the interpersonal sensitivity associated with higher scores on the personality trait of agreeableness are often ineffective in roles requiring cooperation with others (Witt, Burke, Barrick, & Mount, 2002). While research into the interplay of other psychological/situational characteristics suggest that conscientiousness only leads to increased job performance in the presence of ingredients such as a positive psychological climate (Byrne, Stoner, Thompson, & Hochwarter, 2005). Further research into the situational specificity of personality-job performance relations has identified a variety of situational features that operate at task, social, and organisational levels which are relevant to trait expression (Tett & Burnett, 2003; Tett & Guterman, 2000). These and other limitations concerning personality modelling and measurement are likely to be contributing factors to these relatively low estimates of predictive validity (Cervone, 2005; Leising, Erbs, & Fritz, 2010; Read et al., 2010; Srivastava, Guglielmo, & Beer, 2010).

As well as issues of concern relating to the content and simplicity of Big Five models of personality, the design, and theoretical rationale of such models have also been criticised (e.g., Ben-Porath &Waller, 1992a, 1992b; Eysenck, 1992; Hough, 1992; McAdams, 1992; Tellegen, 1993; Waller & Ben-Porath, 1987). Costa and McCrae's popular model built upon the earlier work of Tupes and Christal (1961) and Norman (1963). Costa and McCrae's model has been the target of much of the criticism directed against Big Five personality models, particularly criticism concerning the factor-analytic processes through they were developed (see Boyle, 1989; Boyle et al., 1995; Kline, 1995; Lohman & Rocklin, 1995). Yet the most relevant criticism of the Big Five for this investigation is that its approach to trait classification and broad focus fails to provide the same degree of clarity and predictive value as the facets from which its components are derived (Boyle et al., 1995; Kline, 1995; Mershon & Gorsuch, 1988; Stewart, 1999; Warr et al., 2005). The second-order/global or facet/first-order designation reflects whether or not personality traits are divisible into further measurable units. For example, Big Five openness to experience is a second-order (global) personality trait. This is because it is divisible into the personality facets: empathic, tender-minded, abstract, and radical (Psychometrics Ltd, 2002). Although these facets are defined by very specific characteristics, they are not divisible into further measurable personality dimensions (Cattell, Eber, & Tatsuoka, 1970; Jennrich, & Sampson, 1968). Table 2.2 details the 15FQ+'s facets and global traits.

Table 2.2

15FQ+ Big Five Model of Personality Facet Traits

Global Traits	Facet Traits
Extroversion	Empathic, Enthusiastic, Socially-bold, Group-orientated
Neuroticism	Affected by Feelings, Suspicious, Self-doubting, Tense-driven
Openness	Empathic, Tender-minded, Abstract, Radical
Agreeableness	Low Intellectance, Accommodating, Trusting, Conventional
Conscientiousness	Conscientious, Restrained, Self-disciplined

Note. Derived from information contained within Psychometrics Ltd (2002).

As previously mentioned, the practical utility of personality models in the workplace suggests it is important to also examine the facets comprising the Big Five when trying to understand relationships with job performance. One reason for this is that specific behavioural prediction may be more accurate when focusing upon the finer-grained facets of personality. The rationale for this is straightforward. If one knows someone's position on the behavioural continuum in relation to the Big Five factor of conscientiousness, they can make predictions based upon all the behaviours associated with the facets comprising conscientiousness. For example, this individual is likely to follow rules, have high standards, good attention to detail, exercise control over their behaviour and expression of emotion, and work in a systematic and orderly way. The problem with accuracy is that someone's placement on the conscientiousness continuum could be inflated or deflated due to a greater or lesser propensity in the behaviours measured by any one of these facets. For example, the individual might be very likely to attend to detail and follow rules, but still have emotional outbursts. For this reason, behavioural prediction in the workplace may benefit from not losing sight of facet-level traits (Wood

& Englert, 2009). For a comprehensive examination of the debate over the commercial use of broad or narrow personality instruments see Ones and Viswesvaran (1996), J. Hogan and Roberts (1996) and Schneider, Hough, and Dunnette (1996). See Smith et al. (2003) for a substantial review of the benefits of narrow over broad traits in the area of incremental validity.

The relevance of the claim that facets provide greater clarity than the Big Five was also touched upon in the integrity section. It has been widely theorised, and empirically supported, that the construct domain of integrity tests overlaps with the Big Five traits of conscientiousness, agreeableness, and neuroticism. Which facet traits are involved appears less well established. Ones and Viswesvaran's (1996) paper on this "bandwidth-fidelity" dilemma provides persuasive argument against the purported superiority of narrow over broad personality traits. Yet empirically-based criticisms that second-order models lack the sensitivity of their facet counterparts in predicting relevant outcomes are also convincing. This provides impetus for the inclusion of facet trait analysis in the analyses of relationships among predictors. Section 2.3.3 explores what is known about relationships between personality factors and integrity.

Section 2.3.3: Relationship of Personality Traits to Integrity

A variety of investigations has been undertaken into the relationships between a number of personality traits and integrity test scores (Collins & Schmidt, 1993; Hogan & Hogan, 1989; Hogan & Ones, 1997; Korchkin, 1987; Logan, Koettel, & Moore, 1986; Marcus, Lee, & Ashton, 2007; Murphy & Lee, 1994a; Wanek, 1999; Woolley & Hakstian, 1992). As is consistent with the trajectory of contemporary personality theory, the majority of this research has focused on Big Five personality traits. According to Sackett and Wanek (1996) investigations into the relationship between the Big Five and integrity have pursued two lines of inquiry. The first line of inquiry has attempted to determine the nature of the relationship between integrity tests and the Big Five personality factors such as conscientiousness. The second line of inquiry has attempted to determine whether or not conscientiousness explains the predictive validity of integrity tests. In other words, does partialling out conscientiousness reduce or eliminate the relationship between the criteria of interest and integrity tests? Results pursuing this second line have suggested that although it is correct that the constructs of integrity and conscientiousness are related, it is wrong to think that they are identical or interchangeable (Wanek, Sackett, & Ones, 2003).

Ones and her colleagues have conducted a large-scale data collection and meta-analysis of the relationship between the Big Five and integrity (Ones, 1993; Ones, Schmidt, & Viswesvaran, 1993). Where necessary they converted the personality measures analysed into the Big Five dimensions. Correlations among and between personality and integrity tests were also corrected for range restriction and unreliability through the use of correction factors. The results of these corrections were reported as follows: openness to experience ($r = .09$, $p > .05$), conscientiousness (.39, $p < .05$), neuroticism (-.28, $p < .05$), agreeableness (.34, $p < .05$), and extroversion (.03, $p > .05$). The findings that emerged within this investigation were the same for both overt and personality-based integrity tests. The hypothesised link between integrity tests and conscientiousness was supported in this investigation. Yet integrity tests were also shown to substantially correlate with agreeableness and neuroticism. Ones et al. concluded that as the Big Five traits are not orthogonal, it was reasonable to consider the possibility that the correlations between integrity and agreeableness and neuroticism were simply a reflection of the covariance with conscientiousness, agreeableness, and neuroticism. Ones et al. (1993) tested this possibility by regressing integrity on the relevant Big Five traits. However, their findings and subsequent replications (Ones, Viswesvaran, & Schmidt, 1995; Ones, 2003) have all suggested that conscientiousness, agreeableness, and neuroticism each contribute independently to integrity test scores.

Other research has provided further support for the conclusion that integrity tests tap into conscientiousness, agreeableness, and neuroticism (Byle & Holtgraves, 2008; Hogan & Brinkmeyer, 1997; Ones & Viswesvaran, 1998b). The *Dark Triad* of personality comprises the three independent yet related traits of narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002). A number of investigations into this Dark Triad have found those demonstrating Machiavellianism and subclinical naricissism and psychopathy to score low in agreeableness and conscientiousness, and high in neuroticism (Jakobwitz & Egan, 2006; Paulhus & Williams, 2002). While relationships with agreeableness and conscientiousness appear consistent, others have reported finding negative correlations with neuroticism (Jonason & Webster, 2010).

The importance of conscientiousness, agreeableness, and neuroticism in understanding integrity derives additional support from Digman's (1997) investigation into the superstructure of personality traits. In the process of conducting a factor analysis into the Big Five-based upon multiple data sets, Digman (1997) established the apparent existence of two higher order factors that have come to be known as *Alpha* (α) and *Beta* (β). Alpha factor is a higher-order factor defined by factor loadings from agreeableness, conscientiousness, and neuroticism. Beta factor is another higher-order factor delineated by loadings from extroversion and openness to experience.

Digman's (1997) findings suggest the factors of conscientiousness, agreeableness, and neuroticism may be related to integrity through the latent factor Alpha. The three personality traits from which Digman's Alpha is extracted have also been shown to significantly correlate with situational judgement test performance (Lievens & Coetsier, 2002). Substantial general support for the existence and composition of this Alpha factor has since been provided (DeYoung, 2006; DeYoung, Peterson, & Higgins, 2001; Mount, Barrick, Scullen, & Rounds, 2005; Rushton & Irwing, 2008, 2009). Through multivariate genetic analyses Jang et al. (2006) have also provided evidence for Alpha and Beta factors as stable heuristic devices for integrating personality measurement and developmental theory (*cf.* McCrae et al., 2008). Digman (1997) wrote the following in his discussion of the theoretical meaning of the Factor Alpha:

> Factor α [Alpha] represents the socialisation process itself. From Freud (1930) to Kohut (1977), from Watson (1929) to Skinner (1971), personality theorists of various persuasion have been concerned with the development of *impulse restraint* and *conscience,* and the reduction of *hostility, aggression, and neurotic defense.* From this point of view, factor α [Alpha] is what personality development is all about. Thus, if all proceeds according to society's blueprint, the child develops superego and learns to restrain or redirect id impulses and to discharge aggression in socially approved ways. Failure of socialisation is indicated by neurosis, by deficient superego, or by excessive aggressiveness (pp.1249-1250).

The hypothesised relationship between Digman's Alpha factor and integrity also provides further support for integrity as a measure of contextual performance. Contextual performance is often used synonymously with organisational citizenship behaviours (OCBs) (Organ, 1997; Organ & Ryan, 1995; Podsakoff, MacKenzie, Paine, & Bachrach, 2000). In a review of previous research into OCBs by Podsakoff et al. (2000) it was found that the strongest relationships among dispositional factors and OCBs were those involving facets of the Alpha Factor (conscientiousness, agreeableness, and neuroticism). OCBs were shown by Podsakoff et al. (2000) to be crucial to both performance evaluations and object measures of job performance. OCBs were found to account for an average of 42.9% of variance in performance evaluations across 11 studies reviewed. By contrast, just 9.5% of average variance was accounted for by task performance. OCBs also appear to account for between about 18% and 38% of variance in more objective components of organisational performance, such as financial efficiency indicators and customer service indicators. As well as demonstrating that the same factors underlying integrity test scores underlie contextual performance, Podsakoff et al.'s (2000) review appears to go a long way towards cementing the place of integrity testing in explaining unique variance in the prediction of contextual performance, and thus job performance.

Research within the area of counterproductive workplace behaviour appears to provide further support for this shared underlying relationship for integrity and personality tests with contextual performance or the opposite thereof. While the relationship between integrity and CWB is explicit, a review by Vardi and Weitz (2004) also found a number of personality traits linked to CWB. For example, those with higher scores on neuroticism have been found to perform at lower levels than expected and behave vindictively toward their organisation when the perception of justice is lower or when other more general stressors are involved (Aquino, Lewis, & Bradfield, 1999; Penney & Spector, 2005). Perceptions of justice are influenced by internal factors such as favouritism, inconsistency, discrimination, and external factors such as unfair customer interactions (Rupp & Spencer, 2006; Vardi & Weitz, 2004). Neuroticism also has a clear relationship with other components of CWB such as absenteeism and dissatisfaction (e.g., Elovainio, Kivimaki, Vahtera, Virtanen, and Keltikangas-Jarvinen, 2003; Kisamore, Stone, & Jawahar, 2007; Shi, Lin, L. Wang, & M. Wang, 2009; Vardi & Weitz, 2004). Kotov, Gamez, Schmidt, and Watson (2010) also found neuroticism to significantly correlate with substance abuse disorders ($r = .36$, $p < .05$).

The relationship between personality traits and the integrity construct has also received clarification from work by Wanek, Sackett, and Ones (2003). Wanek et al. undertook a principal component analysis of integrity test items and the relationship of derived composites of the Big Five. This reaffirmed the central role of conscientiousness, agreeableness, and neuroticism in accounting for variance in integrity test scores. It also provided insight into this relationship by revealing the relationships between these factors and the components identified. Conscientiousness and neuroticism were found to significantly correlate across all four of the principal components identified: Antisocial Behaviour (C1), Broad Socialisation (C2), Positive Outlook (C3), and Orderliness/Diligence (C4). On the other hand, while agreeableness was found to correlate with the first three components, it was found to correlate least with the Orderliness/Diligence component (Wanek, Sackett, & Ones, 2003).

The above research suggests integrity tests may measure a higher-order factor of personality. Yet a recent investigation suggests that there may be slightly greater support for overt integrity tests loading on a composite of numerous facets from various domains within the Big Five (Marcus, Hoft, & Riediger, 2006). The relative paucity of findings concerning relationships between facet-level traits and integrity make the generalisation of this finding difficult. Yet relationships among facet-level traits for conscientiousness and integrity have received a reasonable amount of attention.

The results of an examination of investigations into the relationship between integrity and conscientiousness led Sackett and Wanek (1996) to conclude that the key to understanding this relationship lies in the facets of conscientiousness. Sackett and Wanek note that although the Big Five serves as a useful organising device, each of the Big Five traits is multifaceted. This is exemplified by the distinct ways in which the conscientiousness dimension is subdivided by different researchers. For example, Hogan and Ones (1997) identify four subcomponents for conscientiousness: conformity, perseverance, orderliness, and self-control. On the other hand, Costa and McCrae (1985) subdivide conscientiousness into six facets: deliberation, self-discipline, achievement striving, dutifulness, order, and competence. Wanek (1995, 1999) also notes that the relative emphasis placed upon such subcomponents varies across both Big Five and integrity measures. Wanek (1995, 1999) found that Big Five measures of conscientiousness usually place greater emphasis on orderliness, perseverance, and conformity, and less on self-control, which receives the greatest emphasise in integrity tests. Sackett and Wanek (1996) have suggested that it is this differential emphasise on self-control that is responsible for the greater degree of incremental validity of integrity tests over conscientiousness alone. Wanek (1995, 1999) contends that it is the relationship with the conscientiousness component of self-control that explains the integrity-conscientiousness relationship. The 15FQ+ personality measure utilised in chapter three of this dissertation places greater emphasis upon self-control than most measures of conscientiousness.

Woolley and Hakstian (1992) conducted a relatively comprehensive investigation into the relationship between integrity test scores and personality traits. This investigation examined both 16PF and NEO-PI traits. Although a number of significant correlations were detected for 16PF traits, none were particularly strong. The three most noteworthy correlations concerned the factors of ego strength [C] (r = .39, p < .001), superego strength [G] (r = .27, p < .01), and self-sentiment strength [Q3] (r = .34, p < .001). According to Gorsuch and Cattell (1977) these three factors form component parts of a more global factor of personality that reflects "good upbringing" (p. 697). The 15FQ+ dimension of self-control is also comprised of superego strength (G) and self-sentiment strength (Q3) factors comparable to the 16PF. Section 2.4 examines the psychometric construct of intelligence, its relationships with personality traits, and its relationship to job performance.

Section 2.4: Intelligence

"*Intelligence is what intelligence tests test.*"

- Boring (1923, p.36)

This section will explore the applied psychometric construct of intelligence rather than the plethora of other conceptualisations and models of intelligence. This review will focus on a brief examination of hierarchical models of intelligence, such as Cattell's (1957, 1963) model of fluid and crystallised intelligence, Carroll's (1993) three-stratum theory of intelligence, and their recent amalgamation into the Cattell-Horn-Carroll theory of cognitive abilities (McGrew, 2005). It will also discuss the place of ability in predicting future job performance and what is known of relationships between ability, personality, and integrity. The following paragraphs in this introduction briefly outline some alternative models of intelligence. The intention of this is to reinforce the necessarily limited scope of any one particular approach to an investigation of intelligence.

A conceptualisation of intelligence is inextricably bound to the approach within which it is considered. A variety of paradigms has played a significant role within the study of intelligence. The *cognitive-processing* or *developmental* approach to intelligence is one such example (Sternberg et al., 2003). The developmental approach is interested in the method through which people learn to perceive, manipulate, and think about their environment (Piaget, 1973; Sternberg et al., 2003; Vygotsky, 1930-1935/1978). Another example is the *information-processing* approach. The information-processing approach seeks to understand intelligence through the kinds of skills used by people in problem solving and thinking. This generally involves identifying the cognitive processes involved in goal-directed thinking (Sternberg, 1986, 1999, 2000, 2002a, 2003; Sternberg et al., 2003). Considerable attention has also been devoted to explaining the origins of intelligence within the context of evolutionary psychology. Evolutionary psychology conceptualises the mind as an information-processing machine designed by natural selection to solve adaptive problems encountered by our hunter-gatherer ancestors. The primary distinction between competing conceptualisations of intelligence in evolutionary psychology concerns the support of either domain-general or domain-specific perspectives of cognitive architecture (e.g., Cosmides & Tooby, 1997, 2002; Fodor, 1983, 2001; Geary & Huffman, 2002; Girotto & Tentori, 2008; Kanazawa, 2004, 2008; Smith, Borgerhoff Mulder, & Hill, 2001; Wicherts, Borsboom, & Dolan, 2009).

Advances in the technology of neurological assessment coupled with biological explanations for intelligence have led to something of a revival in the popularity of the “speed of processing” type theories originally proposed by Galton (1869; 1885). Biological explanations attempt to determine how the physiology and anatomy or the central nervous system and brain account for intelligent thought (Sternberg, 1990). These biological explanations of intelligence are often referred to as *neural efficiency hypotheses* (Stankov, Boyle, & Cattell, 1995). Neural efficiency hypotheses claim that individual differences in intelligence reflect relative differences in neurological functioning. According to such hypotheses differences in “intelligence” do not reflect a difference in some module or general component of “intelligence.” They instead reflect a difference in the efficiency of brain functions relevant to basic cognitive processes commonly associated with intelligent behaviour, such as communication or pattern recognition.

Another perspective of influence in contemporary conceptualisations of intelligence is *bounded rationality*. Many models of rational inference view the mind as possessing substantial powers of reason, boundless knowledge, and all of eternity with which to make decisions. Bounded rationality instead provides a conceptualisation of the parameters in which humans and other animals make inferences about their world with limited knowledge, time, and computational power (Simon, 1997; Todd & Gigerenzer, 1999, 2000). Barrett (2004) provides an interesting and somewhat sobering commentary on the implications of such a conceptualisation of intelligence for the quantitative assessment of abilities.

The foregoing list of definitions and perspectives concerning intelligence is far from exhaustive, nor are most of these models/perspectives mutually exclusive. It is not the goal of this section to provide such a list, or to analyse, critique, or compare those described. It is not a topic consistent with the aims of this dissertation and is comprehensively addressed in other literature (see Fry, 1984; Neisser et al., 1996; Sternberg, 1982, 1994; Sternberg et al., 2003; Sternberg, & Wagner, 1994). The current investigation’s focus is limited to an applied psychometric conception of intelligence. In spite of the flux of general intelligence conceptualisations, this psychometric approach has continued to function as a benchmark of intelligence testing within I.W.O. psychology (Neisser et al., 1996; Riggio, 2000; Sternberg, Lautrey, & Lubart, 2003). Definitions of intelligence within standard psychometric testing assess one’s capacity for information processing, problem solving, and abstract reasoning. Most intelligence theorists consider such capacities essential components in any conceptualisation of intelligence (Ackerman & Rolfhus, 1999; Feist & Barron, 2003; Horn, 1968; Snyderman & Rothman, 1987). According to consensus, *hierarchical models* provide the best coverage of such capacities (Ackerman & Heggestad, 1997; Carroll, 1982, 1993; Deary, 2000). Hierarchical models are described in section 2.4.1.

Section 2.4.1: The Intelligence Construct

"*…an intelligence test is no more than a sample of the kinds of skills we regard as intelligent.*"
- Vernon (1969, p.11).

Discussions of early developments in the construction of hierarchical models of intelligence generally begin with Charles Spearman's (1927) factor-analytically derived "two-factor theory." Spearman proposed that intelligence is comprised of a general factor (*g*) and a set of specific factors (*s*). According to Spearman *g* is responsible for the positive manifold (observed relations) among intelligence tests. Specific factors are instead responsible for the variance in test scores not explained by *g*. This is because specific factors concern the specific, unique requirements of tasks independent of one's general intelligence. Spearman (1923) divided *g* into the components of *education* and *reproduction.* These components are thought to work together, but to be only moderately correlated and measured separately. Education was conceptualised as the ability to make sense of information encountered. Reproduction was conceptualised as the ability to recall and articulate that information at some future point. These factors are very conceptually similar to Cattell's (1941) model of fluid and crystallised intelligence (Hogan, R., Hogan, J., & Barrett, 2008).

The fluid/crystallised intelligence theory posited that cognitive abilities involving some degree of intelligence are most usefully organised at a general level into two principal classes or dimensions. Fluid intelligence (Gf) was conceptualised as raw intellectual potential and considered dependent upon biological factors such as genotype and central nervous system development (Cattell, 1943b, 1963, 1987c; Horn, 1985; Horn & Cattell, 1966c). The conceptualisation of Gf as innate has achieved considerable popularity among intelligence theorists. However, evidence from behavioural-genetic studies is currently insufficient to substantiate or refute this claim (Richardson, 1998). Crystallised intelligence (Gc) was conceptualised as the primary manifestation of the impact of experiential, educative, and acculturation influences upon intellectual development (Cattell, 1943b, 1963, 1987c;

Horn, 1985; Horn & Cattell, 1966c). Another way to conceptualise this distinction is based upon fluid intelligence reflecting fundamental abilities in reasoning and its related higher mental processes, while crystallised intelligence reflects the size of an individual's knowledge base (Drasgow, 2003). Cattell's fluid and crystallised theory of intelligence has remained one of the most influential within the context of intelligence testing and theorisation (Demetriou & Papdopoulos, 2004; Roberts, Markham, Matthews, & Zeidner, 2005; Sternberg, 2004).

Fluid and crystallised intelligence both rely upon the cognitive processes of perceiving relationships, abstracting, reasoning, problem solving, and concept formation (Horn & Cattell, 1966c). Tests of Gf and Gc can be undertaken via either speed or maximum capability (power) measures based on pictorial-spatial, verbal-semantic, and verbal-symbolic materials (Stankov et al., 1995). For example, Gf is related to speed of reasoning, whereas Gc is related to reading speed. Fluid and crystallised intelligence are also influenced by the relationships among a variety of other factors. Examples of such factors and their relationships include the relationship between one's visual perception/spatial ability and perceptual speed, one's ideational fluency and retrieval capacity, and the relationship between one's reaction time and processing speed (Bates & Shieles, 2003). Horn and Blankson (2005) view Gf and Gc abilities as opposites. Whereas Gc measures indicate the extent to which an individual has incorporated a culture's general and procedural knowledge, Gf measures indicate abilities that depend only minimally on knowledge of the culture.

Fluid intelligence derives its label from its capacity to be channelled into just about any task or activity that requires the exercise of intelligence (Horn, 1977). The construct domain of Gf is comprised of one's general reasoning ability; particularly reasoning that involves figural, symbolic, or nonverbal content (Horn, 1976; Horn & Cattell, 1966c). Fluid intelligence is assessed by tests in which the parameters of elementary relation perceiving and/or correlate-inducing abilities establish one's level of performance (Horn, 1977; Jensen, 1998b). Some characterise the dimension of Gf as nonverbal intelligence (although verbal measures can assess it), or performance IQ. Gf is usually measured by tests involving content such matrices, mazes, letter series, figure classifications, and word groupings (Horn, 1976). A variety of researchers and theorists has attempted to equate Gf with

Spearman's *g* factor (e.g., Cronbach, 1984; Deary, 2000; Gustafsson, 1988; Jensen, 2002; Kyllonen & Christal, 1990). Yet this identification has not been borne out by further investigation (Horn & Blankson, 2005; Kline, 2000; McGrew, 2005; Stankov, 2002). Perhaps the most compelling refutation of the Gf = *g* hypothesis is that presented within Carroll's (1993) seminal work on cognitive ability and its structure. Carroll performed a Schmid-Leiman hierarchical factor analysis and found that an identified second-order Gf remained after the extraction of the third-order *g*.

Crystallised intelligence gets its name from the intellectual abilities it comprises appearing to "crystallise out of experience" (Horn, 1967, p.150). Horn and Cattell (1966c) construe Gc tests as most commonly measuring that aspect of intelligence involving the use of verbal or conceptual knowledge. Gc is seen as concerning one's awareness of terms and concepts pertaining to a wide variety of topics. Gc is generally considered the product of fluid intelligence functioning (Sternberg, 2004). Crystallised intelligence is assessed by items demanding the knowledge and skills it is reasonable to expect of people of the relevant age within a particular culture (Kline, 2000). Crystallised intelligence is normally measured via vocabulary tests and tests that measure general and/or specific knowledge, such as mathematics. Rightly or wrongly, because of its association with the knowledge tested in school aptitude tests etc., Gc is the dimension most usually considered indicative of intelligence (Horn, 1976, 1977). Crystallised abilities are often referred to in attempts to specify what is most important about human intelligence (Horn & Blankson, 2005).

The next major advance in hierarchical models of intelligence was Carroll's (1993) three-stratum theory of intelligence. The three-stratum theory of cognitive ability extends and expands all previous theories of intelligence (Carroll, 1997). Many consider it the most definitive study into the structure of intelligence measured by tests (Grubb et al., 2004; Stankov et al., 1995; Sternberg et al., 2003). McGrew (2005) has even gone so far as to attribute to Carroll the provision of a "Rosetta stone" by which to understand intelligence (p.147). After reanalysing more than 450 data sets covering a full spectrum of tool and participant group diversity, and observing the consistencies across results, Carroll concluded that a hierarchical model of intelligence was the most practically useful and theoretically sound model of intelligence. On the basis of this investigation Carroll proffered a

“Three-Stratum” hierarchy of intelligence. According to this model the stratum of a cognitive ability is assigned on the basis of its perceived breadth or narrowness (Carroll, 1997).

The structure of intelligence within this hierarchy is comparable to a pyramid. Spearman’s *g* is equivalent to the apex of the pyramid (Stratum III). The body of the pyramid (Stratum II) is comprised of eight factors. Each of these factors represent enduring characteristics affecting an individual’s performance in any given area requiring intellectual acumen. The most prevalent of Stratum II factors are fluid and crystallised intelligence. The base of the pyramid, or Stratum I, is comprised of a number of specific factors. Examples of stratum I factors include mathematical reasoning and verbal ability (Carroll, 1997).

Carroll’s (1997) rationale for the creation of his three-stratum theory was to allow the facilitation of a provisional statement about the identification, enumeration, and structuring of the total range of cognitive abilities discovered or known thus far. To this end Carroll’s theory was expected to expand, replace, or supplement previous theories of the structure of cognitive abilities. This included: Thurstone’s (1938) theory of primary mental abilities, Horn and Cattell’s (1966c) Gf-Gc theory, Guilford’s (1967) structure-of-intellect theory, and Wechsler’s (1974) theory of performance and verbal components of intelligence. The next stage in the refinement of a model incorporating both *g* and Gf and Gc was McGrew’s (2005) Cattell-Horn-Carroll (CHC) theory of cognitive abilities.

According to McGrew (2005) the CHC theory of intelligence is “the tent that houses the two most prominent psychometric theoretical models of human cognitive abilities” (p.137). The CHC theory is a hybridisation of Horn and Cattell’s Gf-Gc and Carroll’s three-stratum theory of intelligence. The first published account of the CHC link is contained within Flanagan, McGrew, and Ortiz (2000). This report assigned credit for the CHC link to McGrew’s (1997) attempt to create a single Gf-Gc taxonomy for interpreting and evaluating intelligence test batteries. Neither the Cattell-Horn nor the Carroll model of intelligence was selected over the other in CHC development. The CHC instead synthesised the Carroll and Horn-Cattell Gf-Gc frameworks (McGrew, 1997; 2009). According to McGrew and Flanagan (1997) the empirical evidence strongly suggests that the Horn-Cattell Gf-Gc theory (Horn, 1994; Horn & Noll, 1997) and the three-stratum theory of cognitive abilities (Carroll, 1993, 1997) represent the most well researched and comprehensive frameworks of the structure of intelligence to date (also see Flanagan & McGrew, 1997; McGrew & Flanagan, 1996; Messick, 1992). Although Carroll's model incorporated Gf-Gc factors, it did not assign sufficient weight to the place of these factors in common intelligence testing. The CHC model represents the most accurate model of the factors underlying commonly measured intelligence. This makes CHC taxonomy the obvious "cognitive cornerstone" of modern intelligence theory and test design (Carroll, 1997; Flanagan & McGrew, 1995a, 1998; McGrew, 1997, 2005; Woodcock, 1990).

The validity of CHC over Carrollian second-stratum abilities alone has received considerable support from large and small factorial studies (McGrew, 2005; Shu-chen, Jordanova, & Lindenberger, 1998; Stankov, Seizova-Cajie, & Roberts, 2001). Contemporary CHC modelling has also garnered considerable support from outcome criterion prediction, developmental, heritability, and neurocognitive studies (Floyd, McGrew, Barry, Rafael, & Rogers, 2009; Horn & Noll, 1997).

The CHC model includes a stratum III general intelligence factor. Yet this does not make compulsory its inclusion in research involving the CHC. According to McGrew (2005) researchers employing the CHC model need to themselves decide whether *g* is to be included in the application of the CHC in their research. An evaluation of the relative merits of the *g* versus no-*g* approaches to intelligence theorisation are beyond the scope of this dissertation; at least four substantial books or major papers have been exclusively devoted to the topic (Brand, 1996; Jensen, 1998b; Nyborg, 2003; Sternberg & Grigorenko, 2002) and no consensus has been reached. Regardless of the reality of *g,* there is no doubt that Spearman's conceptualisation has had, and continues to have a profound impact upon intelligence theorisation and test construction. A considerable amount of empirical evidence suggests that general intelligence is among the most dominant and enduring factors connected with environmental adaptation; occupational success; and physical propensity and morbidity (McGrew & Flanagan, 1997; McDermott, Fantuzzo, & Glutting, 1990). The use of *g* in psychometrics remains such an emotive topic that authors such as Deary (2000, 2001) claim theories failing to account for *g* are "pseudo-science." Figure 2.3 represents the CHC model of ability.

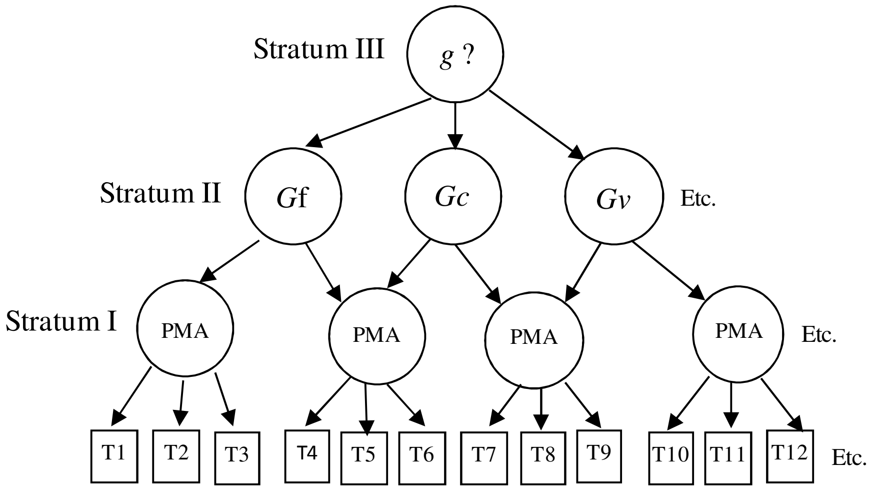


Figure 2.3. CHC Hierarchical Model of Intelligence (adapted from McGrew, 2005). Circles represent latent factors (PMA = primary mental ability). Squares represent manifest measures (T = test).

This dissertation will adopt a conservative approach to the inclusion of *g*. A general intelligence factor will be included within analysis, but will assume a secondary role to Gf and Gc within the discussion. Evidence that secondary abilities such as Gf and Gc make important contributions to understanding cognitive performance above and beyond that accounted for by *g* is substantial (McGrew & Flanagan, 1997; McGrew, 2005, 2009). Yet the inclusion of *g* remains important from a practical perspective. Composite intelligence tests involving *g* comprise several different abilities. This means they are often better predictors of overall performance ratings than measures of Gc or Gf alone (Horn, 1994). This is important from a validity perspective, yet similar arguments apply to those made earlier made regarding a focus on the Big Five or their facets.

Section 2.4.2: Intelligence Test Prediction of Job Performance

The appropriate use of intelligence tests within the workplace can greatly increase the accuracy and value of personnel selection (Hawk, 1986; Schmidt, Ones, & Hunter, 1992). Schmidt and Hunter (1998) demonstrated that intelligence tests were one of the most valid methods of predicting future job performance across a wide variety of jobs. Hunter (1980) demonstrated that a measure of general intelligence was able to predict employee performance in all kinds of jobs regardless of their nature or complexity. In a U.S.A. Department of Labor commissioned meta-analysis (Hunter, 1980; Hunter &

Hunter, 1984) trainability and job performance was measured for over 32,000 employees in 515 different civilian jobs. The job categories within this meta-analysis ranged from unskilled and skilled blue-collar jobs to midlevel white-collar administrative jobs and professional managerial jobs. Test validity for the tests varied from .28 for completely unskilled jobs to .58 for professional managerial jobs. The validity of .51 demonstrated among middle-complexity jobs, which equates to approximately 60% of the jobs in the U.S.A. economy, is a reasonable and commonly used estimate (Grubb et al., 2004).

General mental ability measures are at present the best possible and most widely available method of predicting job-related learning. A variety of studies have indicated that psychometric measures of intelligence are the most accurate predictor of employees' ability to acquire on the job knowledge (Schmidt & Hunter, 1992; Schmidt, Hunter, & Outerbridge, 1986). They are also the best predictors of employees' ability to perform in job training programmes (Hunter, 1986; Hunter & Hunter, 1984; Ones et al., 2005a; Ree & Earles, 1992). Another major benefit of the use of ability tests in selection decisions is their price. Ability tests have the lowest application cost of selection processes that are suitable for all categories of job. A professionally developed ability test, applicable to all jobs, costs just a few dollars per test administration (Grubb et al., 2004).

The possible economic gains an organisation can accrue via the utilisation of ability measures in selection decisions are enormous (Grubb et al., 2004). Schmidt and Hunter (1998) have summarised methods for measuring employee performance as a function of dollar value or productivity relative to an average level of performance. Higher performing employees are conservatively thought to do 40% more work than their lower performing colleagues (Schmidt & Hunter, 1983; Schmidt, Hunter, McKenzie, & Muldrow, 1979; Schmidt, Mack, & Hunter, 1984). This difference represents considerable variance in both productivity and dollar value to New Zealand organisations.

New Zealand organisations constantly strive to achieve maximisation of cost effectiveness in personnel selection in terms of predictive validity and selection utility. This has led to a growth in the ubiquity of cognitive ability test use in employment selection (Guenole et al., 2003). A survey of selection practices within New Zealand across 100 randomly selected organisations and 30 recruitment firms established that almost two thirds of recruitment firms used cognitive ability measures in selection procedures (Ryan et al., 1999). This survey also found that nearly one-half of all organisations examined employed intelligence tests in the selection of managerial personnel. This is more than two times the amount utilised for this purpose a decade ago (Taylor, Mills, & O'Driscoll, 1993). Another recent survey into intelligence test use in personnel selection examined the comparative prevalence of such use across 18 countries. The prevalence of intelligence test use in selection decisions in New Zealand was greater than all but three of those countries examined (Ryan et al., 1999). The use of intelligence tests in selection decisions continues to grow. Section 2.4.3 details the relationship between intelligence and integrity.

Section 2.4.3: Relationship of Intelligence to Integrity

The predictive validity of intelligence tests are to some extent influenced by the complexity of the job to which they are applied (Hunter & Hunter, 1984). For general intelligence tests, the validity of the tests increases as the level of job complexity increases. However, there has been concern that the opposite effect may hold true for integrity test validities. Ones et al. (1993) hypothesised that as the level of job complexity increases, a systematic decline in the estimated validity of integrity tests would occur. It was thought this would be due to relatively intelligent applicants more accurately judging which responses would be perceived as unfavourable and manipulating their answers accordingly. It was also proposed that this relative intelligence would pose greater problems for detecting deviant behaviours in high-complexity jobs. Such concerns have been historically tested through examinations of the relationship between intelligence and integrity test scores.

For instance, Gough (1972) investigated the relationship between the Personnel Reaction Blank and the College Vocabulary Test for participant groups of 125 males and 68 females. Gough reported correlations of .01 and -.05 for these participant groups respectively. Jones and Terris (1983) investigated the relationship between the London House PSI attitude and admissions scales and the 16PF B scale, which is often considered a brief measure of general intelligence. Jones and Terris found nonsignificant correlations of -.03 for theft attitudes and -.02 for theft admissions against the 16PF B scale for a participant group of 104 convenience store applicants. In a later study Werner, Jones, and Steffy (1989) found a similar absence of any significant correlation between the 16PF B scale and theft proneness. J. Hogan and R. Hogan (1989) also investigated the relationship between integrity and intelligence tests. The Hogans correlated the Hogan Reliability Scale with the quantitative and verbal scales of the Armed Services Vocational Aptitude Battery and found correlations of .07 and -.09 respectively for an unspecified participant group of 97. A review of such early investigations into the relationship between intelligence and integrity by Sackett et al. (1989) found near-zero correlations between integrity tests and measures of cognitive ability.

Such research suggests no relationship of significance between integrity and intelligence. Yet there are methodological concerns that limit the confidence in such a conclusion. Jones and Terris's (1983) and Werner et al.'s (1989) use of the 16PF B scale as an intelligence measure is highly problematic. The 16PF B scale was originally conceptualised by Cattell to encapsulate those aspects of the personality indicative of intellect. However, it is not comprised of items comparable to an intelligence test, but rather items determining whether attitudes are consistent with intellectual interests. Such attitudes are not sufficiently correlated with standard intelligence tests to be considered a reasonable measure to employ as an intelligence proxy. There are also problems with the investigations of Gough (1972) and J. Hogan and R. Hogan (1989). The previous subsections of this intelligence section explicated the importance of the fluid/crystallised differentiation in intelligence testing. Without making this differentiation it is not possible to determine the influence of upbringing and education on scores. Without such a determination relationships between variables such as integrity and intelligence can be obscured. Yet neither investigation employed intelligence measures capable of fully illuminating any such relationships – significant or not.

Subsequent meta-analytic reviews conducted by Ones (1993) and Ones et al. (1993a) provided further support for the conclusion that intelligence tests scores do not substantially correlate with integrity scores. Ones et al. acquired a corrected correlation of .02 from 106 investigations between seven integrity measures and 12 intelligence tests. A corrected correlation of -.01 was derived for the relationship between overt ($k = 29$) integrity measures and intelligence. Such results indicate that there are no significant or noteworthy differences in relationships between integrity test types and intelligence. This investigation has continued to serve as the justification for subsequent claims that the relationship between intelligence and integrity is nonsignificant (Sackett & Wanek, 1996; Wanek, 1999).

More recent reviews of investigations into integrity and ability have reiterated the general lack of any significant relationship (Berry et al., 2007; Cullen & Sackett, 2004). Furthermore, investigations that have found significant relationships have suggested these are likely to be small. For example, Mehrabian (2000) examined the relationship between intelligence and integrity with a participant group of university students and employees (107 men and 195 women) aged 17-46 years ($M = 20.03$, $SD = 4.04$). Mehrabian tested this relationship via the Wonderlic Personnel Test measure of intelligence. This is a widely used timed test of general intelligence. He correlated Wonderlic scores with scores from *The Integrity Scale*. This is an overt measure of integrity that assesses an individual's honesty, trustworthiness, reliability, and dependability. Mehrabian found that integrity had a positive correlation of .15 ($p < .05$) with intelligence, which is significant if small. Other investigations have also found similarly significant, yet small correlations with ability and the kinds of CWBs predicted by integrity tests (Dilchert, Ones, Davis, & Rostow, 2007; Marcus et al., 2009; Roberts, Harms, Caspi, & Moffitt, 2007).

There is also potential benefit in combining more general personality scales with intelligence in selection systems. Section 2.4.4 examines research into the relationship between intelligence and personality.

Section 2.4.4: Relationship of Intelligence to Personality Traits

Ackerman and Heggestad (1997) conducted a meta-analysis based upon 135 separate studies into the relationship between intelligence and personality. Ackerman and Heggestad's findings suggested that intelligence is positively correlated with personality traits falling under the rubric of positive affect. Openness to experience and extroversion are examples of such positive traits. Ackerman and

Heggestad's findings also suggested that intelligence is negatively correlated with traits representing negative affect. Neuroticism is an example of such negative traits. Austin et al.'s (2002) investigation into the relationship between intelligence and personality in four large datasets provided further support for this finding. A variety of other investigations have also suggested a negative relationship between neuroticism and intelligence (Hembree, 1988; Kyllonen, 1997; Seipp, 1991) or ability-related criteria such as academic performance (Chamorro-Premuzic & Furnham, 2003).

Relationships between facets of neuroticism and ability tasks have also been the subject of some investigation. Depression appears to relate to cognitive impairment in functions such as reading comprehension and problem solving (Hartlage, Alloy, Vasquez, & Dykman, 1993). A tendency for increased aggression and delinquency to be associated with low intelligence was also revealed in research on the effects of anger (Zeidner, 1995).

The most extensively researched personality trait in contemporary investigations, and that most strongly correlated with intelligence is openness to experience (Zeidner & Matthews, 2000). Most such research has supported a strong positive relationship between openness to experience and intelligence (Ackerman & Goff, 1994; Costa, Fozard, McCrae, & Bosśe, 1976; Holland, Dollinger, Holland, & MacDonald, 1995; Hough, 2001; McCrae & Costa, 1985, etc.). Investigations into the relationship between intelligence and personality that distinguish between Gc and Gf have generally found openness more strongly correlated with Gc than Gf (Bates & Shieles, 2003; Brand, 1994; Goff & Ackerman, 1992; Kanfer et al., 1995; Van der Zee, Zaal, & Piekstra, 2003). Goff and Ackerman (1992) found a correlation of $r = .40$ between Gc and openness. They suggested that this relationship was likely to be a consequence of those high in openness being more motivated to engage in intellectual activities. Moutafi et al. (2003) suggest that further support for the link between openness and intelligence is derived from studies of authoritarianism. These studies have found authoritarianism negatively correlated with both openness ($r = -.57$) (Trapnell, 1994) and intelligence (ranging from r –.20 to –.50) (Zeidner & Matthews, 2000). The relationship between openness and intelligence is consistent with the conceptual similarity of these constructs (Barrick & Mount, 1991;

Eysenck, 1992; McCrae & Costa, 1987). The fact that few or no significant relationships are reported between integrity and these factors is also consistent with the openness-intelligence link.

Studies into the relationship between intelligence and extroversion have been somewhat less consistent. Ackerman and Heggestad's (1997) meta-analysis suggested a significant, albeit very small, positive correlation between extroversion and intelligence ($r = .08$). However, other investigations into this relationship have found significant negative correlations between extroversion and intelligence (Austin et al., 2002; Furnham et al., 1998a). Moutafi et al. (2003) suggest that the contradictory nature of such findings is likely due to the use of tests measuring different aspects of intelligence within these investigations. Zeidner (1995) suggested that extroverts are advantaged in performance tasks (automatic motor sequences) and introverts are advantaged in respect to verbal tasks (associative learning). Moutafi et al.'s (2003) and Zeidner's (1995) suggestions both re-emphasise the importance of examining sub-facets of the Big Five and the hierarchical components of intelligence (*g,* Gf, and Gc*).*

Agreeableness and conscientiousness are generally the least correlated with intelligence of the Big Five personality factors (Moutafi et al., 2003). Research by Ackerman and Heggestad (1997), and Kyllonen (1997) found near zero correlations between agreeableness and *g*. These researchers reported similarly insignificant relationships between conscientiousness and *g*. However, recent research has reported significant negative correlations between conscientiousness and intelligence (Moutafi et al., 2003, 2004; Wood, 2004; Wood & Englert, 2009). Moutafi et al. (2004) found Big Five conscientiousness to be negatively correlated with Gf ($r = -.21$, $p < .01$), but not Gc ($r = -.09$, $p > .05$). They also found Gf and Gc respectively significantly negatively correlated with the 15FQ facets of conscientious ($r = -.26$, $p < .001$; $r = -.29$, $p < .001$) and disciplined ($r = -.16$, $p < .05$; $r = -.16$, $p < .05$). Wood (2004) and Wood and Englert (2009) also reported significant negative relationships between intelligence and conscientiousness. They found negative relationships between conscientious facets and Gf and Gc for both the 15FQ (conscientious and disciplined) and OPP (conformity and detail-Conscious) personality measures. However, not all conscientious facets were found to correlate significantly with Gf and Gc for either the 15FQ (tense-driven and restrained) or OPP (phlegmatic).

This provides further support for including facet-level traits in any analysis of predictor relationships with personality.

Unlike findings by Moutafi et al. (2004), Wood and Englert (2009) found Gc to correlate more strongly with conscientious facets than Gf across all measures. Wood and Englert explained this disparity on the basis of the fundamental difference between the populations employed in these respective investigations (high-achievers versus average people who have not graduated from university). Findings that the relationships between conscientiousness and intelligence differ across conscientiousness facets and intelligence components again supports the importance of making such fine-grained distinctions when conducting the research necessary to understand the relationships amongst these traits.

As well as such empirical relationships between intelligence and personality test scores, some interesting conceptualisations of the relationships between personality and intelligence have been proffered. In one such example good judgment is proposed to occur as a function of relationships between intelligence and personality (Hogan, R., Hogan, J., & Barrett, 2008).

Section 2.5 examines the use of response time information as an indicator of response style and time management. It also briefly examines the response as a parameter of ability and moderating influence on ability scores.

Section 2.5: Response Time (RT)

Previous research has exhibited considerable interest in the amount of time it takes a test taker to respond to a given stimulus, particularly in the areas of psychophysics and cognitive psychology. This response time (RT) research has traditionally focused on inferring mental process organisation from RT distributions required in different mental process components (e.g., Luce, 1986). This focus has resulted in a number of interesting outcomes, including the *speed-accuracy trade-off function* (SATF), which describes how response time influences a respondent's accuracy in performing certain tasks (Luce, 1986; Roskam, 1997). Jensen (1998) provides a comprehensive review of such response time related research and theories.

The contemporary ubiquity of computer-based testing has now seen educational measurement and more general psychometrics pay considerable attention to the relationship of response time to test taking strategies, item difficulty, and examinee ability (e.g., Jansen, 1997; Parshall, Mittelholtz, & Miller, 1994; Schnipke & Scrams, 1997; Scrams & Schnipke, 1997; Thissen, 1983). Schnipke and Scrams (1998) provide an overview of this early literature. This overview suggests that there is a strong yet complicated relationship between response time and accuracy, which is influenced by test attributes such as whether or not there is a time limit on responding. With the increasing use of computer-based testing the need for more research into how RT data can improve measurement quality will only increase (Wang & Hanson, 2005). Response time within this context is generally defined as the amount of time elapsed between question presentation on the computer screen and an examinee's response to that question (Verbic, 2010). Section 2.5.1 briefly outlines the use of RT as a measure of response style.

Section 2.5.1: Response Time as Response Style

The history of response time research in psychometrics has spanned a variety of areas. One is the use of RT to detect distortion in personality assessments. Response latency and Item Response Theory (IRT) approaches both attempt to detect faking through understanding differences in item-response processes. Item response theory-based approaches to detecting faking have met with some success (Ferrando & Anguiano-Carrasco, 2009; Stark, Chernyehenko, Chan, Lee, & Drasgow, 2001). However, these methods have not been consistent in demonstrating their effectiveness over and above that of existing methods (Ferrando & Chico, 2001; Zickar & Drasgow, 1996). Latency measurement is based upon the idea that candidates attempting to distort items are likely to take relatively shorter or longer periods of time to respond to those items (Holden & Kroner, 1992). Although debate continues concerning the efficacy of the latency approach (e.g., Vasilopoulos, Reilly, & Leaman, 2000), some relatively compelling results have been presented in its favour (Holden & Kroner, 1992; Tetrick, 1989; Walczyk, Roper, Seemann, & Humphrey, 2003).

Response time data has also been used to detect preknowledge of ability items within an item pool and the potential differential speededness of tests (van der Linden & Guo, 2008; van der Linden & van Krimpen-Stoop, 2003). Preknowledge of ability items in the pool is likely to result in unexpectedly correct responses in unexpectedly short response times. Differential speededness of tests reflects the concern that items requiring more time for a correct response may be overrepresented in CAT for some examinees. The use of RT to establish how much time is needed to answer items of varying difficulty may allow CAT administration time-constraints to be fitted to item requirements. This would seem an important goal considering the implicit assumption that IRT modelled tests are not administered under speeded conditions (van Breukelen, 2005). Research into the use of RT for detecting aberrant behaviour is nascent, but is also showing promise in the detection of warming-up effect for the examinee, fatigue built up toward the end of the test, and misunderstanding of the instructions (van der Linden & Van Krimpen-Stoop, 2003).

Research has also suggested RT can be used in personality assessments to increase the accuracy of assessment whilst reducing the number of items required (Ferrando & Lorrenzo-Seva, 2007). Furthermore, response times have been involved in attempts to increase reliability through an understanding of the relationship between speededness and response style. It is generally thought that increasing the number of items in an assessment will increase reliability, but RT research suggests that such an increase may be of limited value on timed assessments where test takers start guessing when running out of time to answer questions towards the end of assessments (Attali, 2005). This is discussed further in section 2.5.2.

Section 2.5.2: Response Time and Time Management

The only time-limit related information respondents typically get during an ability assessment is the overall minutes and seconds remaining. This can be problematic from a time management perspective as respondents can only guess how much time current and upcoming items will require and should be allotted (van der Linden, 2009). The way in which respondents' manage their time reflects the pacing strategy adopted (Millman, 1969). It is possible that providing respondents with information on time restrictions *per item* will facilitate more consistent engagement in solution-based behaviour across all items. As previously noted, respondents often switch from solution-based behaviour to rapid guessing for items towards the end of an assessment with an overall time limit (Wise & DeMars, 2006). The use of per item timing is also consistent with van Breukelen's (2005) recommendations and gathering information at this level will optimise subsequent calculations of test time requirements.

The adoption of a good pacing strategy is an important aspect of sound test-taking strategy (Millman, 1969). Examples of pacing strategies include working rapidly at the beginning of timed tests and omitting items appearing excessively difficult in order to focus on those more solvable (Wang & Zhang, 2006). While this may be a good strategy for Classical Test Theory (CTT) designed assessments, omitting difficult items on CAT assessments may result in an artificially deflated ability estimate (Chang & Ying, 2008). Pacing strategies adopted are also closely related to examinee effort, which is in turn closely related to whether or not the testing occurs in high or low stakes circumstances (Wise & DeMars, 2006).

The time required to complete items is an important consideration in test designs where an emphasis is placed upon skill rather than speed. Power tests are traditionally defined as those that have items ranging in difficulty. These tests are intended to measure how accurately respondents can answer questions. In contrast, pure speed tests have very easy items and strict time limits; the goal being to measure how quickly respondents can answer. Despite this theoretical distinction, most ability assessments used in the workplace have items of varying difficulty, but also use time limits as a matter of administrative convenience (Schnipke & Scrams, 1997; Scrams & Schnipke, 1997). In order to increase the capacity for time management within the second study of this dissertation and further clarify the relationships between predictor scores and RT, respondents will be provided with time remaining information per item rather than as an overall limitation. Section 2.5.3 briefly reviews theorised relationships between RT and ability or personality test outcomes.

Section 2.5.3: Response Time Relationship with Individual Differences

Considerable recent research has examined RT as a form of collateral information in IRT-based estimations of response and accompanying item parameters. Van der Linden, Entink, and Fox (2010) suggest the incorporation of RTs can enhance the accuracy of respondent ability estimates and improve the selection of items within CAT administration. Despite such findings testing agencies remain hesitant to report scores that are calculated from RT and other information that does not reflect response accuracy (van der Linden, 2008). This testing agency reluctance only increases the relevance of calls to replicate and extend findings on RT and item parameters (e.g., Scrams & Schnipke, 1997; Wang and Hanson, 2005). Furthermore, such resistance does not impact upon the important role RT information can have in designing tests and determining realistic time limits for tests and batteries (Entink, Fox, & van der Linden, 2009; van der Linden, Entink, & Fox, 2010). The use of RT information in this way is likely to be acceptable to testing agencies and stands to advance the accuracy of ability score calculations considerably (van der Linden, 2008).

The use of RT as an ability parameter is based upon the idea that the speed at which test takers complete items and the overall assessment is an important criterion of ability (Glickman, Gray, & Morales, 2005; Scrams & Schnipke, 1997). This is consistent with the historical view that speed of problem solving is an important aspect of intelligence (Thorndike, 1913; McFarland, 1930). The majority of support for this view comes from investigations involving processing speed on elementary cognitive tasks (ECTs), not RT performance on actual ability items (e.g., Grudnik & Kranzler, 2001; Jensen, 1998; Petrill & Deary, 2001). One reason for this focus on ECT derived RTs is to avoid confounds with other factors that influence a respondents' test taking speed. Jensen (1998) differentiates what he calls *test-taking speed* and information-processing speed. He attributes information-processing speed to observed relationships between RT and ability and aligns test-taking speed with personality tendencies. Research focusing on ECT derived RTs has provided some interesting information regarding the relationship between information-processing and ability. For example, Neubauer (1997) reported a significant correlation between intelligence and speed of information processing (r = .30). According to such findings more intelligence people have more efficient brain and broader central nervous systems (Rindermann & Neubauer, 2001). As noted earlier, this is generally referred to as the neural efficiency hypothesis (Jensen, 1982).

Some research has examined RT as actual test-taking speed. Yet the relationship between RT and examinee ability has not always been significant, strong, or in the direction expected. Scrams and Schnipke (1997) found that examinee ability only accounted for 8.80% of log response time variance in a verbal test and 6.31% in a reasoning test. Neubauer (1990) reported no significant correlation between intelligence estimates based on Raven's Progressive Matrices and response times. On this basis it is not surprising that some research has treated speed as a nuisance factor. Unfortunately such research is almost always conducted in circumstances far removed from those encountered by real-world candidates. Real-world candidates most regularly complete ability assessments under tight time constraints, while participants in RT research tend to complete assessments without time limits or with very liberal time limits (Jensen, 1998). This again reinforces the need to collect and test RT relationships in settings more consistent with those encountered by job applicants.

As mentioned in section 2.5.2, time management and associated pacing strategies can have a significant impact on test outcomes. In line with this, test takers who adopt a slow and careful approach to answering questions are likely to be penalised by not finishing the assessment (Bridgeman, 1980). Such a slow and careful approach is consistent with the characteristics of conscientiousness, and although yet to be tested, this has led some researchers to question whether generally conscientious respondents are penalised in standardised timed testing contexts (Jensen, 1998; Powers & Kaufman, 2004). Evidence of this would provide an interesting and important alternative to the ICT theory's idea that less intelligent individuals compensate through the adoption of conscientious behaviour (Wood & Englert, 2009).

Relationships between ability and personality have been observed to differ within timed or untimed testing conditions. For example, Rawlings and Carnie (1989) found a significant relationship between extroversion and performance on the WAIS information subtest to be consistent across timed and untimed conditions. Yet other significant relationships were unique within timed or untimed conditions. In one such example introverts were found to be superior to extroverts on the digit span task, but only within untimed conditions. Other research has also suggested that personality traits might interact with response times to influence ability outcomes. Bates and Eysenck (1993) attributed a *lack* of personality interactions with response times to the absence of opportunities to employ response strategies within an untimed research design. Other researchers have found response latency on a neuroticism scale negatively associated with ability (Furnham et al., 1998a). The possible existence of a negative relationship between ability response times and neuroticism is consistent with the previously mentioned neural efficiency hypothesis. This hypothesis suggests respondents experiencing relatively high levels of anxiety engage in substantially more task irrelevant processing (i.e., worry) (Moutafi et al., 2005). This is consistent with the idea that neuroticism affects ability test *performance* rather than actual ability (Chamorro-Premuzic & Furnham, 2004; Zeidner & Matthews, 2000). Moutafi et al.'s (2005) finding that neuroticism can explain more variance in fluid intelligence indices than those of crystallised intelligence is consistent with this explanation as fluid intelligence items require greater processing of resources than crystallised items.

Chapter 2 Summary

A key question within this dissertation was which psychometric predictors of job performance should be the focus of investigation. This chapter identified integrity, conscientiousness, neuroticism, and fluid and crystallised intelligence as the predictors of relevance. This choice was based upon likely job performance coverage, potential gains in incremental validity, and reductions in adverse impact. A key consideration concerning incremental validity and adverse impact reduction was identified as correlations amongst predictors.

Section 2.2 suggested integrity tests are an effective tool for minimising the impact of dishonest and deviant behaviour in the workplace. Section 2.3 identified the Big Five traits of conscientiousness and neuroticism as important predictors of job performance. While observed relationships were reported for conscientiousness or neuroticism and integrity, these Big Five traits were still expected to add incremental validity to integrity tests. Furthermore, section 2.1.5 suggested that the addition of conscientiousness, neuroticism, and integrity to ability assessments was likely to reduce the potential or extend of adverse impact experienced by some ethnic minorities.

Section 2.4 examined the psychometric concept of intelligence. Intelligence tests were reported to be the most accurate psychometric predictor of future job performance. Section 2.4 also detailed a number of relationships amongst intelligence, integrity, and more general personality traits. Highlights of these findings included the lack of significant relationship between integrity and ability, neuroticism and conscientiousness being related to ability, and Big Five personality trait facets varying in their relationship with ability.

Section 2.5 briefly outlined the use of response time within an assessment context. This included information exploring how the length of time taken to respond to items within an assessment could be used as an indication of attempted distortion or the resonance of item content for a respondent. The use of response time to further understand and control time-management/pacing strategies was also discussed. Finally, the possibility that the length of time spent on assessments may vary as a function of individual differences was explored. Some researchers suggested response time was one criteria of ability. Some also suggested that differences in the personality traits of conscientiousness and neuroticism could negatively impact upon ability test performance during timed administrations. The next chapter details the research objectives to be pursued within this dissertation.

Chapter 3: Research Objectives

"…in the progress of investigation, we should proceed from known facts to what is unknown."

- Lavoisier (1789, p.1).

This chapter outlines the broad research objectives to be explored in this dissertation. The research objectives are derived from the literature review and not intended to encompass all possible issues or questions. The objectives seek to clarify relationships between integrity, conscientiousness, neuroticism, and fluid and crystallise ability. The dissertation also presents a way to measure those constructs that have the potential to increase predictive validity, reduce adverse impact, and incorporate contemporary test-design strategies.

Section 3.1: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability.

The first research objective was prompted by the need to better understand relationships among integrity, conscientiousness, neuroticism, and ability. Examining these relationships may clarify the construct domains of these predictors. This may assist making more accurate predictions of future job performance. It may also facilitate the previously mentioned calculations concerning potential reductions in adverse impact by combining predictors.

Research reported in Section 2.3.3 suggested that as integrity increases, so does conscientiousness (e.g., Ones, 1993; Ones et al., 1993a; Sackett & Wanek, 1996), but neuroticism falls as integrity increases (Ones et al., 1995; Ones, 2003). Theorists also suggest that the key to understanding these relationships may lie at the facet level (Sackett & Wanek, 1996; Wanek, 1995, 1999). Looking for clarification of relationships at the facet level is also consistent with findings reported in Section 2.4.4, which suggested that relationships between personality traits and ability were not consistent at the facet level (e.g., Moutafi et al., 2003, 2004; Wood, 2004; Wood & Englert, 2009).

Research Objective 1: To investigate whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. Furthermore, to explore variations among any such differences at the facet level.

Section 3.2: Adverse Impact Risk and Reduction for those identifying as Māori

As seen in section 2.1.5, when ability assessments are used for selection, ethnic minority groups are sometimes hired at lower ratios than majority group members (e.g., Campion et al., 2001; Grubb et al., 2004; Roth et al., 2001). Previous research conducted within the context of New Zealand workplace testing has suggested that those identifying as Māori may also experience such adverse impact (Guenole et al., 2003; Packman et al., 2005).

Adverse impact may reduce job opportunities for Māori and make the workforce less ethnically diverse (Sackett & Wilk, 1994; Sackett et al., 2001). It is also inconsistent with the spirit of the State Sector Act (1988) and the principle of partnership within the Treaty of Waitangi (1975). One way to reduce adverse impact is to create test batteries that include assessments without average scores displaying significant ethnic differences (Sackett & Ellingson, 1997). This dissertation will investigate whether or not Māori are at risk of adverse impact, and whether this risk can be reduced by combining ability tests with other assessments.

Research Objective 2: To investigate whether there is a difference in personality or ability test scores between those identifying as Māori and non-Māori. Then to explore whether any such differences could be reduced by combining ability tests with other assessments.

Section 3.3: Response Time

The present dissertation also asked if psychometric tests could be improved by measuring response time. The section examining the relationship between response time and individual differences (2.3.5) suggested that the length of time taken by a respondent to answer ability items could be used as one estimate of their ability (Glickman, Gray, & Morales, 2005; Scrams & Schnipke, 1997). The most consistent support for this view comes from investigations involving processing speed on simple tasks rather than response times for items of varying difficulty. This research suggests higher ability respondents complete simple cognitive tasks more quickly than lower ability individuals. (e.g., Grudnik & Kranzler, 2001; Petrill & Deary, 2001). Other research suggests that respondents take longer to answer more difficult items (Neubauer, 1990; Verbic, 2010).

Researchers within Section 2.3.5 also theorised that the slow and careful approach characteristic of conscientiousness may mean that conscientious respondents score lower on ability results because they have less opportunity to display solution-seeking behaviour across questions during timed testing (Jensen, 1998; Powers & Kaufman, 2004). Furthermore, if response times increase as neuroticism increases, this may suggest more neurotic respondents score less well on ability assessments due to engaging in substantially more task irrelevant processing (i.e., worry) (Moutafi et al., 2005). This is consistent with the ideal that neuroticism affects ability test *performance* rather than actual ability (Chamorro-Premuzic & Furnham, 2004; Zeidner & Matthews, 2000). It is also consistent with research that has found response latency on a neuroticism scale negatively associated with ability (Furnham et al., 1998a).

Research Objective 3: To investigate whether the length of time taken to answer ability assessments is related to respondent ability or item difficulty. Moreover, to explore whether differences in conscientiousness or neuroticism scores meant respondents take longer to answer ability assessments.

Chapter 4: Study 1

The purpose of this chapter is to detail the investigation undertaken into relationships among integrity, intelligence, conscientiousness, and neuroticism. Understanding such relationships is important for construct clarity and having the information necessary to perform a number of calculations relating to the second research objective of this dissertation. The calculations that will be performed in this study examine adverse impact, but having information on predictor interactions will also help facilitate future calculations of incremental validity. This investigation used a data set provided by Occupational Psychology and Research Associates (OPRA) Group Ltd. OPRA are an organisational psychological consulting firm who specialise in the use of psychometrics in selection systems. The first sections of this chapter detail the research design and materials used. This includes a brief outline of the general and psychometric properties of measures used. It also includes an examination of how the research design and construct operationalisation of tests fits with the goals of the dissertation. Following sections detail the procedure used to conduct this investigation and its results. These results are then discussed in light of previous findings, composite design implications, and wider contributions to the field in general. The research objectives investigated in this chapter are:

Research Objective 1: To investigate whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. Furthermore, to explore variations among any such differences at the level of facet or ability subcomponent.

Research Objective 2: To investigate whether there is a difference in personality or ability test scores between those identifying as Māori and non-Māori. Then to explore whether any such differences could be reduced by combining ability tests with other assessments.

Section 4.1: Method

This investigation employs a cross-sectional correlational design. It involves the within-subjects examination of 20 factors. Sixteen of these factors are personality traits, three are cognitive ability factors, and one is an integrity factor. The correlational paradigm is consistent with the aims of this study. The factors of interest within this dissertation are participants' personality traits, intelligence, and integrity. The correlational design also provides a method of hypothesis testing with appropriate ecological validity concerning both testing context and participant group composition (Bordens & Abbott, 1996).

The correlational design examines relationships among all variables, including relationships that are not specifically the focus of research objectives. This facilitates the identification of any previously unknown significant relationships among variables. The identification of such relationships is important for knowledge of covariance and potential directions for future research (Bordens & Abbott, 1996). Exploring research objectives while also identifying previously unknown relationships is consistent with Tukey's (1977) emphasis on research designs combining both exploratory and confirmatory goals.

The correlational design's lack of causal attribution for relationships and ability to analyse more than two variables is also consistent with the multivariate nature of this investigation. A multivariate approach to investigating individual differences is widely supported as behavioural tendencies often appear interrelated (see for reviews, Boyle, 1991; Nesselroade & Cattell, 1988). The multivariate approach is also extensively employed within the Cattellian School (Boyle & Cattell, 1984; Stankov, 1980, 1987, 1989; Stankov & Chen, 1988). This is a relevant consideration as the personality and intelligence components of this investigation have their origins with the Cattellian School. The personality measure used is a derivative of Cattell's Sixteen Personality Factor Questionnaire (16PF), and the cognitive ability measure utilised serves as an operationalisation of Cattell's Fluid and Crystallised model of intelligence (section 4.1.2). The following section provides information on the participant group used.

Section 4.1.1: Participants (Study 1)

Two hundred and eleven job applicants applying for pre-managerial roles participated in this investigation. Demographic information for these participants is detailed in Table 4.1. The participant group was 60% female and 40% male. A slight majority of participants identified as NZ European/Pakeha (35%), followed by Māori (22%), Pacific Peoples (16%), and Asian (10%). The majority of participants reported possessing a qualification falling into the "other" category (33%). The next most common educational attainment reported was a Bachelor's Degree (22%). OPRA's Directors provided access to archival cross-sectional data sets associated with these participants. These data sets were acquired by OPRA within the context of its commercial practice. The data were obtained from individuals who had completed the 15FQ+, SSI, and GRT2 measures in New Zealand between April 2003 and December 2008. These assessments were completed within the context of application for employment. Data were acquired from OPRA's Auckland, Wellington, and Christchurch offices. The data acquisition procedures are reported after the measures section.

Table 4.1

Demographic Statistics, OPRA Investigation participant group

	Frequency	%
Gender		
Male	85	40
Female	126	60
Ethnicity		
NZ European/Pakeha	75	35
Māori	46	22
Pacific Peoples	34	16
Asian	21	10
Missing	35	17
Age		
17-20	12	6
21-25	27	13
26-30	38	18
31-35	25	12
36-40	24	12
41-45	39	18
46-50	19	9
51-55	12	6
56-60	4	2
61-65	1	1
Missing	10	5
Education		
No formal education	12	6
Secondary School Qualification	21	11
Bachelor's Degree	45	22
Higher University Degrees	26	13
Other	68	33
Missing	31	15

Note. N = 211. Percentages may not add up to 100 because of rounding. Original ethnicity categories labelled NZ Europeans/Pakeha as "Caucasian" and Pacific Peoples as "Pacific Islander; these were changed to reflect modern conceptualisations.

Section 4.1.2: Measures (Study 1)

This section provides information on the assessments utilised within this investigation. This information relates to the operationalisation of their constructs, the components of each measure, and their psychometric properties. As consistent with the standard criteria for assessment legitimacy, psychometric information focuses on reliability and validity (e.g., AERA/APA/NCME, 1999; Campbell & Fiske, 1959; Kline, 2000; Landy, 1986; Nunnally & Bernstein, 1994; Spector, 1981). The measures examined are:

- The General Reasoning Test Battery (GRT2)
- The Fifteen-Factor Personality Questionnaire (15FQ+)
- Stanton Survey of Integrity (SSI)

Section 4.1.2.1: The General Reasoning Test Battery (GRT2)

The General Reasoning Test Battery (GRT2) was designed to serve as a composite psychometric measure of intelligence within the workplace (Budd, 1993). The GRT2 comprises the commonly tested and important verbal, numerical, and abstract reasoning ability subtests (Ackerman, Kanfer, & Goff, 1995; Ackerman & Rolfhus, 1999; Heim, 1970).

The GRT2's composition is consistent with this investigation's Cattell-Horn-Carroll (CHC) conceptualisation of intelligence. Verbal, numerical, and abstract reasoning represent first stratum components of the CHC model and are subcomponents of the second-stratum factors of fluid and crystallised intelligence. Fluid and crystallised intelligence are in turn subcomponents of stratum-three's general intelligence factor (Carroll, 1993). As well as functioning as measures of the second-stratum factors of fluid and crystallised intelligence, the GRT2 components of verbal, numerical, and abstract reasoning can be summed to provide a general intelligence score (Cowan, 2005; Jensen, 1998b).

The verbal reasoning component assesses one's knowledge of words and ability to use words in a rational way. This latter requirement involves accurately identifying logical relationships between words and using these relationships to draw appropriate inferences and conclusions (Budd, 1993). This verbal reasoning subtest primarily loads on crystallised intelligence (Carroll, 1993).

The abstract reasoning component assesses the ability to identify logical relationships between abstract geometric patterns and spatial relationships. This type of abstract reasoning assesses fluid intelligence. This means the GRT2's abstract reasoning component loads upon reasoning ability that is not greatly affected by educational and cultural experience (Budd, 1993).

The numerical reasoning component assesses the capacity to use numbers in a reasonable way, correctly identifying logical relationships between numbers and drawing appropriate conclusions and inferences upon this basis. The numerical reasoning component of these measures is less clear-cut than verbal or abstract reasoning in its loading on crystallised or fluid intelligence. Numerical reasoning assessments are generally comprised of relatively novel and thus reasonable assessors of fluid intelligence. Yet a certain minimal level of arithmetic attainment (Gc) is required because numbers cannot be manipulated except by mathematic rules (Kline, 2000). A basic level of arithmetic attainment is also required to accurately answer GRT2 numerical reasoning items (Budd, 1993).

General Properties

The GRT2 is a power test targeted at the general population. The technical manual suggests no significant difference between versions administered in paper and pencil or computer format, or scored manually or electronically (Budd, 1993). This is consistent with other findings concerning the psychometric equivalence of computer and booklet forms of tests (Finger & Ones, 1999). Those undertaking the GRT2 are asked to work as quickly and accurately as possible and are given the opportunity to try example questions prior to undertaking the test proper.

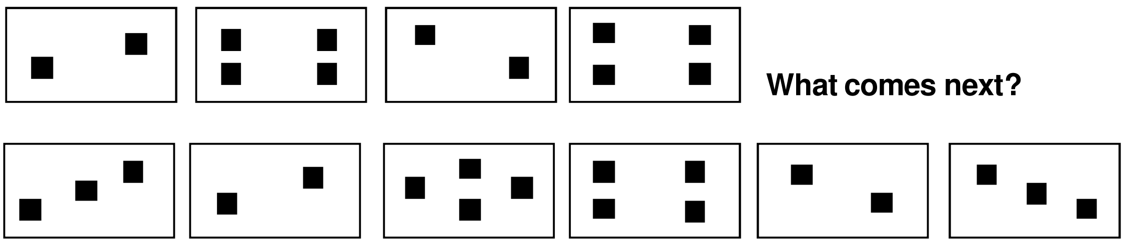
The GRT2's verbal component has 35 items with six response options and has an eight minute time limit. An example item follows:

Sick means the same as?
1. Labour 2. Ill 3. Healthy 4. Disease 5. Evil 6. Rest

The numerical component has 25 items with six response options and a time limit of 10 minutes. An example item follows:

Which of the following is the odd one out?
1. 2/4   2. 6/8   3. 3/6   4. 4/9   5. 4/8   6. 2/6

The abstract component has 25 items with six response options and a time limit of 10 minutes. An example item follows:



Psychometric Properties

Information pertaining to the reliability and validity of the GRT2 follows. This information provides support for this measure's capacity to assess verbal, numerical, and abstract reasoning ability.

Reliability Evidence

A test-retest investigation after a two week interval carried out by the test supplier found acceptable levels of consistency across testing occasions: verbal (r = .81), numerical (r = .84), and abstract (r = .78) (N = 54). The GRT2 technical manual also details Alpha coefficients for its three subscales as: verbal (r = .83), numerical (r = .84), and abstract reasoning (r = .83) (N = 135). That each of these reliability coefficients exceeds 0.8 supports the GRT2's internal consistency (Budd, 1993; Psytech, 2009).

Validity Evidence

Construct validity support provided by the supplier for the GRT2 comes from comparisons with other ability assessments. The first reports correlations between GRT2 and The Group Test of General

Reasoning Skills (AH3) subscales, and total scale scores for both measures (N = 81). The AH3 is designed to measure the cognitive ability of the general population (Heim, Watt, & Simmonds, 1974). Correlations were as follows: verbal (r = .63), numerical (r = .76), abstract (r = .56), and total (r = .82) (Budd, 1993; Psytech, 2009). The strength of these correlations indicates that the GRT2 and AH3 are measuring similar constructs. Yet correlations between the GRT2 and AH3 fail to provide clear support for the discriminant validity of GRT2 inferences. This is because the correlations between each of the GRT2 subscales and their respective AH3 counterparts (e.g., GRT2 Verbal with the AH3 Verbal) are not significantly greater than correlations across other presumably less related subscales (e.g., GRT2 Abstract with AH3 Verbal [r = .61]).

The GRT2 technical manual also details correlations between the GRT2 and other measures of comparable constructs, namely the Technical Test Battery (TTB2) and the Clerical Test Battery (CTB2). General Reasoning Test subscales have modest correlations with the TTB2 (N = 94) ranging from .40 to .47. When the GRT2's verbal reasoning subtest was compared to the CTB2 (N = 54) the strongest correlation was not as expected with the Spelling measure (SP2) (r = .34), but instead with the Office Arithmetic scale (NA2) (r = .51). Nevertheless, subsequent examination of the NA2 does reveal a reasonably high verbal problem-solving element, which may explain this relationship and provides further support for the validity of inferences based upon GRT2 scores.

A number of criterion-related validity studies for the GRT2 are also reported. In a concurrent investigation the GRT2 significantly correlated with the numerical (NR2, r = .29; AR2, r = .31) and software (NR2, r = .32; AR2, r = .28) related skill areas of 118 retail bankers. In another study involving service engineers (N = 46), rated performance was found to correlate strongly with the GRT2 subscale of verbal reasoning (r = .46), and moderately with the abstract reasoning subscale (r = .28) (Budd, 1993). Another study found the job performance of experienced newspaper printers (N = 70) was correlated with abstract (r = .36), verbal (r = .26), and numerical reasoning (r = .28). A further investigation examining the aggregate examination performance of 100 trainee Financial Consultants found correlations with numerical (r = .46) and abstract reasoning (r = .44). These investigations and others provide evidence for criterion-related validity in support of GRT2 inferences (Budd, 1993; Psytech, 2009).

Section 4.1.2.2: The Fifteen Factor Questionnaire – Second Edition (15FQ+)

The Fifteen Factor Questionnaire Plus (15FQ+) is a personality measure designed for use within the workplace. The 15FQ+ comprises 16 facets reducible to five second order traits consistent with the Big Five. Table 2.1 and Table 4.2 respectively detail the 15FQ+'s Big Five and facet-level personality traits. The 15FQ+ also contains a number of response style, or impression management indicators. The social desirability scale assesses a tendency to present oneself in an unrealistically positive manner. It is a dedicated scale that is independent of the primary personality factors. The infrequency scale measures the extent to which someone has attended diligently to the questionnaire and responded in a consistent way. A high score on this scale suggests a respondent may have rushed through the assessment or failed to fully understand questions. The so called central tendency scale assesses the extent to which people have been prepared to answer the questionnaire decisively rather than providing non-committal responses (Psychometrics Ltd, 2002).

General Properties

The 15FQ+ does not have a time limit for completion. Most people complete it in about 25-30 minutes and respondents are instructed not to spend too much time pondering over each question. Respondents are also instructed to answer all questions while avoiding middle (uncertain) answers and being as honest and truthful as possible. Participants are also provided with an opportunity to undertake example questions in preparation of the test proper. The 15FQ+ has 200 items with three response options. An example item follows:

I like to go to the cinema

1. True 2. ? 3. False

Table 4.2

15FQ+ Facet Traits

Trait	Low Score Description	High Score Description
FA	Cool Reserved: Lacking empathy, Distant, Impersonal, Detached	Empathic: Personable, Friendly, Caring, Warm-hearted, Participating
FB	Low Intellectance: Lacking confidence in own intellectual ability	High Intellectance: Confident in own intellectual ability
FC	Affected by Feelings: Emotional, Labile, Moody	Emotionally Stable: Mature, Phlegmatic, Calm
FE	Accommodating: Mild, Passive, Deferential	Dominant: Assertive, Aggressive, Competitive
FF	Sober Serious: Restrained, Cautious, Taciturn	Enthusiastic: Lively, Cheerful, Happy-go-lucky
FG	Expedient: Disregarding of rules and obligations, Spontaneous	Conscientious: Persevering, Detail-conscious, Dutiful
FH	Retiring: Timid, Socially anxious, Hesitant in social settings, shy	Socially Bold: Venturesome, Socially confident, Talkative
FI	Hard-headed: Utilitarian, Lacking aesthetic sensitivity, Unsentimental	Tender-minded: Sensitive, Sentimental, Aesthetic
FL	Trusting: Accepting, Unsuspicious	Suspicious: Cynical, Sceptical, Doubting
FM	Concrete: Realistic, Solution-focused, Practical, Down-to-earth	Abstract: Imaginative, Absent-minded, Absorbed in thought, Impractical
FN	Direct: Genuine, Open, Artless, Direct, Straightforward	Restrained: Diplomatic. Socially aware, Socially astute, Shrewd
FO	Confident: Self-assured, Secure, Unworried, Guilt-free	Self-doubting: Apprehensive, Insecure, Worrying
FQ1	Conventional: Conservative, Traditional	Radical: Experimenting, Unconventional, Open to change
FQ2	Group-orientated: Group dependent, Sociable, a “Joiner”	Self-sufficient: Solitary, Individualistic, Self-reliant
FQ3	Informal: Undisciplined, Lax, Follows own urges	Self-disciplined: Compulsive, Exacting willpower, Fastidious
FQ4	Composed: Placid, Relaxed, Patient	Tense-driven: Impatient, Low frustration tolerance

Note. Adapted from Psychometrics Ltd (2002).

Psychometric Properties

Information pertaining to the 15FQ+’s reliability and validity is presented below. This information is derived from the test supplier and supports the legitimacy of this measure’s facet and secondary (global) traits.

Reliability Evidence

Test-retest coefficients between .77 and .89 (N = 87) supports the 15FQ+’s reliability. The 15FQ+ technical manual also presents Cronbach’s α coefficients for both student (N = 183) and professional (N = 325) participant groups. The student participant group shows a range of coefficients from .74 to .84. The professional participant group shows a coefficient range from .76 to .83 (Psychometrics Ltd, 2002). The strength of these coefficients supports the 15FQ+'s internal consistency without raising concerns that scores are so high they may simply reflect repeated coverage of very similar items (i.e., *bloated specifics*).

Validity Evidence

The construct validity of 15FQ+ inferences was assessed via comparisons with a variety of other personality measures, including: the original 15FQ, the 16PF–4 (Form A), and 16PF-5 (Psychometric Ltd, 2002). The construct validity of 15FQ+ global factor scores was also gauged via comparisons with NEO PI-R, 16PF-4, and 16PF-5 global factors. The 16PF assessments reflect Cattell's (1957) original personality trait conceptualisations. The NEO PI-R assesses McCrae and Costa's (1987) Big Five. These measures have strong psychometric properties and serve as excellent benchmarks upon which the performance of other personality tests can be determined.

Ten of the sixteen reported corrected correlations between the 15FQ+ and 15FQ very strongly support the validity of 15FQ+ personality dimensions. These ten correlations range from .93 to 1. Of the remaining six dimensions all but two correlate substantially with their respective 15FQ factors (.73 – .84). The two factors with the lowest correlations are empathic (FA) (.43) and tense-driven (FQ4) (.40). The 15FQ+'s empathic factor focuses on warm-hearted empathic concern for, and interest in, other people. The original 15FQ empathic factor instead focuses on sociability and interpersonal warmth. This difference is likely to account for the modest correlation between these factors. The modest correlation between the 15FQ+ and 15FQ factors of tense-driven (FQ4) is equally explicable. Whereas the 15FQ+ factor of tense-driven assesses a tense, competitive, hostile interpersonal attitude, the 15FQ dimension instead assesses emotional tension and neuroticism (Psychometrics Ltd, 2002). Correlations between the 15FQ+ and the two versions of the 16PF also support the validity of 15FQ+ dimensions. Twelve of the correlations reported between the 15FQ+ and 16PF (Form A) range from .70 to 1. The remaining four range from .37 to .65. Correlations between the 15FQ+ and 16PF – 5 follow a similar pattern. Thirteen of these correlations ranged between .70 and 1. The remaining correlations range from .31 to .56 (Psychometrics Ltd, 2002).

The strength of the correlations between the 15FQ+ and two 16PF versions supports 15FQ+ construct validity. Nevertheless, it is notable that 15FQ+ factor conscientious (FG) correlated most strongly with 16PF – 5 dimension of self-disciplined (FQ3), while 15FQ+ factor of self-disciplined (FQ3) correlates most strongly with 16PF – 5 dimension of conscientiousness (FG). According to the 15FQ+ technical test manual (Psychometrics Ltd, 2002) this is because the meaning of these two factors has been reversed in the fifth edition of the 16PF.

Relationship between 15FQ+ global factors and 16PF-5 global factors strongly support 15FQ+ global validity ($N = 85$). Substantial correlations between the 15FQ+ global traits of extroversion ($r = .88$), agreeableness ($r = .81$), and neuroticism ($r = .87$) indicate that these broad personality constructs are measuring comparable constructs. Although slightly lower, correlations between the global factors of openness to experience ($r = .65$) and conscientious ($r = .79$) remain high and provide further support for 15FQ+ global validity. This validity is also supported by the very low correlations between 15FQ+ global factors and divergent 16PF –5 global factors (Psychometric Ltd, 2002).

Further support for the validity of inferences made on the basis of 15FQ+ global factor scores is derived from comparisons with NEO PI – R factors ($N = 60$). Correlations provided in the 15FQ+'s technical test manual are as follows: extroversion ($r = .74$), neuroticism ($r = .77$), openness ($r = .66$), agreeableness ($r = .61$), and conscientious ($r = .67$). These correlations support the broad equivalence of the 15FQ+ global factors and the Big Five personality factors defined by Costa and McCrae (1987).

Section 4.1.2.3: Stanton Survey of Integrity (SSI)

The SSI was designed to serve as a self-report psychometric measure of integrity within the workplace (Harris & Gentry, 1992). The SSI comprises components concerning *work related theft*, *theft unrelated to work*, and *company policy violations*. This is consistent with the continued demonstration of the importance of these facets within integrity testing (O'Bannon et al., 1989; Sackett & Wanek, 1996; Wanek, Sackett, & Ones, 2003).

The work related theft component of the SSI deals with a person's self-reported behaviour in the workplace. It assesses involvement in theft at previous jobs or associating with others engaged in deviant behaviours. It also determines an applicant's views concerning theft in the workplace. Items assessing the theft unrelated to work component of the SSI explore a respondent's deviant behaviour outside of the workplace. Such behaviour is believed to generalise from outside of work to a respondent's place of employment. The SSI hypothesises that a lenient attitude toward criminal behaviour outside the workplace is one of several possible indicators of subsequent CWBs. The company policy violations component examines the likelihood of a respondent violating broader company policies. Examples of such violations include: abusing sick time, arriving late for work, or leaving work early (Harris & Gentry, 1992).

General Properties

The SSI is administered during the pre-employment phase of a job application with the express purpose of reducing a company's losses due to employee CWBs (Harris, 1987). The administration mode in which the SSI is taken does not appreciably affect the scores acquired (Viswesvaran & Ones, 1997). This is consistent with other findings suggesting no significant difference in norm scores for test versions administered or scored manually or electronically (Finger & Ones, 1999; Psychometric Ltd., 2002).

There is no time limit within which to complete the SSI. However, it takes most people about 10-20 minutes to complete and respondents are instructed to go with the first response that comes to mind. Those undertaking this measure are further instructed to: work quickly and accurately, choose only one answer, and make sure they answer all questions. There are 83 items. Sixty-two items are forced-choice and 21 are multiple choice. An example of these types of questions and their response options are respectively:

Have most workers taken things from work for which they could be convicted if caught?
1. Yes 2. No

Which of these categories do you feel best describes your own honesty?
1. Extremely honest 2. Almost totally honest 3. Slightly dishonest

Psychometric Properties

Information pertaining to the reliability and validity of the SSI is provided below. This information is in part provided by the test supplier and supports its psychometric legitimacy.

Reliability Evidence

In a 1994 test of the SSI's test-retest reliability a participant group of 316 applicants were compared across a time range of between 1 and 365 days, with an average of 47.3 days (Stanton, 2005). The results of this investigation showed a correlation of .7. Viswesvaran and Ones (1997) also assessed the SSI's test-retest reliability as satisfactory.

Harris and Gentry (1992) also conducted a test to assess the internal consistency of the Stanton Survey after its last revision. This investigation indicated that the internal consistency of the SSI ranged from .91 to .93. In a subsequent review of a variety of studies Viswesvaran and Ones (1997) reported high internal consistency reliabilities for the SSI (.90s depending on the specific internal consistency index used).

Validity Evidence

During the Survey's revision evidence in support of construct validity was acquired via comparisons between the attitude scale of the SSI on the one hand, and the Guilford-Zimmerman Temperament Survey (GZTS) and Personality Research – Form E (PRF – E) on the other (Harris & Gentry, 1992). Correlations with the GZTS suggest that applicants with low percentile scores on the SSI attitude scale are likely to be impulsive, self-centred, introverted, hostile, overly critical, easily excited, and resistant to authority. The following PRF-E scales were positively correlated with the SSI attitude scale: autonomy, aggression, defendence, impulsivity, and play. The following PRF-E scales were found to negatively correlate with the SSI attitude scale: achievement, endurance, cognitive structure, nurturance, order, harm-avoidance, and understanding. Applicants with low SSI percentile scores tend to be argumentative, nonconforming, hot-tempered, defensive, suspicious, and unpredictable. They are also prone to being impatient, foolhardy, overly adventuresome, and immediate gratification seeking. These correlations with GZTS and PRF-E scales support SSI validity.

During this research into the revision of the Stanton Survey the New Edition was correlated with the original version of the SSI. The correlation coefficient between these two versions of the measure was .97. The strength of this correlation means the original and new Survey can be viewed as parallel forms (Harris & Gentry, 1992).

SSI construct validity has also received support by correlating SSI scores with other measures of individual differences. Convergent validity was supported in studies assessing relationships between workplace relevant Big Five personality characteristics. The highest of these correlations was with conscientiousness, the next highest with agreeableness, and the lowest with neuroticism (Ones, 1993; Ones et al., 1995). Those exploring the SSI's discriminant validity evidence found negligible correlations with the Wonderlic Personnel Tests (ability tests) (Viswesvaran & Ones, 1997).

In response to criticism that integrity tests have lacked predictive validity studies, Dillon and Harris (1988) demonstrated the predictive validity of the SSI in a year-long study. Figures for inventory shrinkage were compared among three new retail stores from the same organisation. The stores were similar in physical size and layout. Store 1 administered the SSI to applicants, but did not use findings in making hiring decisions. Store 2 administered the SSI to applicants and used the results in making hiring decisions. Store 3 served as the control store and used their "normal" selection processes in hiring decisions. Turnover and inventory shortage in these stores was tracked for one year. At the end of this period it was found that Store 2 had the lowest rate of turnover and lowest inventory shortage of the three stores. The criterion-related validity of the SSI has also found substantial support in a variety of studies examining the concurrent validity of this measure (Viswesvaran & Ones, 1997).

Section 4.1.3: Procedures (Study 1)

All those involved in the acquisition of these data had undergone training in the administration of psychometric tests consistent with British Psychological Society's standardised procedural guidelines for psychometric test administration. These guidelines required testing to have occurred in quiet, well-lit, and ventilated rooms with sufficient space to work. All materials necessary for a test's completion were provided to participants prior to the introduction of the test. If paper-and-pencil administration was undertaken these materials included test booklets, written response forms, demographic information sheets, and writing implements. OPRA test administrators then followed five standardised steps:

1. Participants were welcomed and informed of the purpose of the tests in general terms (it was part of the selection process etc.). Participants were then informed of the benefits they could expect from completing these measures.
2. At this stage informed consent was sought from participants. This entailed gaining permission for using their results and any demographic information provided for future, anonymous statistical/research purposes. No pressure was applied to participants in order to solicit consent (only those who consented to their results and other information being available for statistical and research purposes have been included in this investigation).
3. Instructions on test completion from the administrator's manual were then followed. Confirmation from participants that they understood what is required was then sought. If necessary further clarification was then provided, without going beyond that indicated as permissible in the manual.
4. Participants were then allowed to complete the test example questions prior to starting the test proper. The test administrator then checked that these example questions were completed correctly. The need to work as quickly as possible to complete each test was then communicated. Any further questions a participant may have had were answered at this time

– as long as an answer did not provide undue advantage. Participants were then left to complete the assessments in peace.

5. Participants were then thanked for completing assessments and informed of the next stage in the selection procedure. This included information on subsequent contact and timelines concerning the opportunity to receive feedback on their assessment results.

Categorical (nominal) participant data was then edited and recoded into a numerical format for the purpose of analysis. The recoded data included participant demographics for gender, ethnicity, and education. The statistical analyses necessary to achieve the goals of this research were then undertaken. These analyses are detailed in the following section.

Section 4.1.4: Analyses (Study 1)

Consistent with the research objectives of this investigation, statistical analyses explored the degree of shared variance among variables. The strength of association measure used was Pearson-Product Moment correlation coefficients (i.e., Person's r). Pearson's r is the most appropriate index of shared variance with continuous and normally distributed data (Ferguson, 2009). A number of significant relationships were found among the factors examined. How practically important statistically significant relationships are considered to be depends partly on whose guidelines are employed. Cohen (1992) recommends that effect sizes of approximately .1 are small, .3 are moderate, and .5 are large. On the other hand, Ferguson (2009) recommends the small label is applied to r effect sizes of approximately .2, medium to .5 or thereabouts, and strong to effect sizes of approximately .8. Ferguson's more conservative recommendations are based upon a desire to ensure greater practical significance of results. This investigation will discuss results consistent with Cohen's recommendations, but results meeting Ferguson's (2009) recommendations will be given more weight.

Another consideration within this investigation was the appropriate p-level to employ. The arbitrary nature of the standard .05 p-level regularly used by researchers in social science has led to serious criticisms (Kirk, 1996; Rosnow & Rosenthal, 1989). This investigation instead determined the most relevant and appropriate probability values to focus on via a Bonferroni adjustment. The

Bonferroni adjustment is a calculation applied to make it more "difficult" for any one test to be statistically significant when performing multiple statistical significance tests on the same data (Maxwell & Delaney, 2004; Winer, Brown, & Michels, 1991). A Bonferroni adjustment for the calculations contained within Table 4.4 means that any test that results in a *p*-value of less than .004 is considered statistically significant. A Bonferroni adjustment was also performed for the calculations performed in Table 4.5. This adjustment returned a probability value of $p < .002$. As with effect size considerations, results meeting the less stringent .05 criterion will be discussed, but those meeting the more stringent Bonferroni-based requirements will be given more weight.

Power was another consideration of note for ensuring statistical confidence in the results of computations (Weinfurt, 2003). The highly restrictive α probability values derived from the earlier Bonferroni adjustments would normally entail a high risk of β errors (Cohen, 1988). Fortunately this investigation was able to minimise this risk by successfully retaining the widely recommended power figure of .8 (Spicer, 2005). An earlier power analysis had identified the minimal participant group size (N = 132) necessary for this investigation to achieve this level of power for an effect size of .2. This participant group size was exceeded in the majority of analyses (N=211).

Cohen's discrimination index (*d*) was calculated to examine potential differences in mean scale scores between those identifying as Māori and the broader participant group. This is consistent with the second research objective of this dissertation and the intention to facilitate subsequent calculations for potential adverse impact reduction. The calculation was based on Sackett and Ellingson's (1997) formula. The significance of differences was calculated by an independent measures *t*-test.

Calculations were also made to determine the significance of differences among observed correlations between the scales examined. These were based on Streiger's (1980) recommendations for both large participant group sizes and Type I error control.

Multiple regressions were also performed to establish how much unique variance in integrity was explained by the Big Five and facet-level traits examined. These regressions all utilised integrity as the criterion variable. Criterion choice was based upon previously hypothesised relationships and further explication of correlations observed in analyses of coefficients. Predictor variables included the Big Five factors of conscientiousness, neuroticism, and agreeableness. Predictor combinations including conscientiousness and neuroticism facets were also regressed. Other predictors combined and examined were fluid, crystallised, and general intelligence.

The rationale for regression analysis was the additional information such analyses would provide about the relationships among predictors and integrity. Of particular interest were differences in how much combinations of Big Five personality traits versus their facets contributed to the prediction of integrity. Multiple regressions were performed because the information provided concerns the combined variance explained by a group of predictors rather than the individual relationships previously detailed (Cohen, Cohen, West, & Aiken, 2003).

Section 4.2: Results (Study 1)

This section details analyses utilising the StatSoft® statistical software package *Statistica 7*. A number of statistically significant relationships of small to medium strength were observed amongst the factors examined. The strength and significance of relationships found between personality traits and integrity or cognitive ability varied depending upon whether Big Five or facet traits were examined. Correlation matrices are detailed for Big Five and facet traits in Table 4.4 and Table 4.5 respectively. Table 4.6 details the amount of unique variance in integrity accounted for by different levels and combinations of conscientiousness and neuroticism.

Section 4.2.1: Descriptive Statistics

Table 4.3 details participants' descriptive statistics. This includes information on ranges, means, and standard deviations for participants' scores on the 15FQ+, SSI, and GRT2. The statistics contained within Table 4.3 are reasonably consistent with those of a more general New Zealand pre-management working participant group. The mean scores and standard deviations for the distributer provided GRT2 scores in the 2010 general New Zealand norm are: abstract (M = 16.64, SD = 4.86, N = 41765); numerical (M = 14.01, SD = 5.67, N = 40574); verbal (M = 22.40, SD = 5.68, N = 41182); and *g* (M = 52.81, SD = 14.27). The mean scores and standard deviations for SSI scores in the 2010 general New Zealand norm are: SSI Score (M = 40.07, SD = 11.83, N = 1042). These means and standard deviations for the GRT2 and SSI are within a single standard deviation of difference from those reported for participants within this investigation. The mean scores and standard deviations for 15FQ+ scores in the more general participant group (N = 13200) are also consistent with those reported in Table 4.3.

The consistency between Table 4.3's descriptive statistics and those of a more general New Zealand working participant group supports the potential generalisability of this investigation's findings. Participants within this investigation were non-randomly selected. This means it is not justifiable to make generalised statements on the basis of its findings. However, the similarity between this participant group's descriptive statistics and those of more general and potentially representative samples suggest findings may be replicable with a sample randomly drawn from the target population of applicants for roles within organisations that use psychometrics. This similarity and use of archival cross-sectional data sets obtained with a real-life testing context support the ecological validity of this investigation.

Table 4.3

Descriptive Statistics for 15FQ+, SSI, and GRT2 Dimensions

Factors	Valid N	Mean	Min.	Max.	Std.Dev.
15FQ+ Facet Traits					
Empathic	211	19.71	10	24	3.22
High Intellectance	211	18.27	3	24	4.45
Emotionally Stable	211	16.46	5	24	4.46
Dominant	211	13.60	2	22	4.22
Enthusiastic	211	15.75	2	24	4.66
Conscientious	211	18.13	4	24	4.66
Socially Bold	211	13.45	0	24	5.60
Tender-minded	211	15.39	4	24	4.81
Suspicious	211	6.61	0	20	4.48
Abstract	211	11.31	3	22	4.15
Restrained	211	18.39	6	24	4.00
Apprehensive	211	13.04	0	24	5.56
Radical	211	7.97	0	20	4.23
Self-sufficient	211	7.12	0	22	4.00
Self-disciplined	211	18.56	4	24	4.06
Tense-driven	211	9.35	0	23	5.06
Social desirability	211	9.98	1	15	3.21
15FQ+ Big Five					
Extroversion	211	15.00	2.44	24.36	4.17
Neuroticism	211	5.75	-4.81	17.51	4.46
Openness	211	15.98	5.53	26.23	3.95
Agreeableness	211	-3.95	-14.24	4.35	3.55
High Self-control	211	22.37	9.21	29.79	4.21
GRT2 Factors					
Abstract Reasoning	211	14.21	2	25	5.84
Numerical Reasoning	211	11.70	1	25	6.07
Verbal Reasoning	211	19.61	6	35	6.92
Total (g)	211	45.52	14	85	16.90
Stanton Survey					
Score	211	39.50	9	61	11.05
Credibility	211	7.49	3	10	1.80

Note. As with all tables contained within this investigation, only the "high-scoring" label of the bipolar personality dimension is reported.

The on the basis of checks undertaken the data appeared to met the required statistical assumptions to perform multiple regressions and correlations (Cohen et al., 2003). More specifically, the data appeared normally distributed in histograms; displayed linearity and homoscedasticity; and measurement appeared reliable (Osborne & Waters, 2002). The following section explores the first research objective.

Section 4.2.2: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability

This section presents results relevant to the present dissertation's first research objective. More specifically, it investigates whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. Furthermore, it explores variations among any such differences at the level of facet or ability subcomponent.

Section 4.2.2.1: Correlation Coefficients

Table 4.4's results suggest integrity is positively correlated with Big Five conscientiousness (r = .18, p<. 05). Although this relationship was significant, it did not meet the Bonferroni threshold and the strength of the relationship was small. Table 4.4 also suggests integrity is negatively correlated with Big Five neuroticism (r = - .41, p<. 004). This relationship was of moderate strength. Neuroticism also has small to moderate significant correlations with the indices of fluid (r = -.32, p<. 004), crystallised (r = -.24, p<. 004), and general ability(r = -.31, p<. 004).

Table 4.4

Correlation Matrix for 15FQ+ Global factors, the GRT2, and the SSI

	Abstract	Numerical	Verbal	*g*	Extraversion	Neuroticism	Open	Agreeable	Self-Control	SSI Score
Abstract	**1.00**	**0.67****	**0.68****	**0.87****	**0.17***	**-0.32****	0.11	0.04	-0.12	0.10
Numerical	**0.67****	1.00	**0.74****	**0.89****	0.04	**-0.29****	0.11	-0.05	-0.13	**0.16***
Verbal	**0.68****	**0.74****	1.00	**0.91****	0.08	**-0.24****	**0.18***	-0.07	**-0.14***	**0.14***
g	**0.87****	**0.89****	**0.91****	1.00	0.10	**-0.31****	**0.15***	-0.03	**-0.14***	**0.15***
Extraversion	**0.17***	0.04	0.08	0.10	1.00	**-0.22****	**0.29****	**-0.25****	**-0.17***	0.09
Neuroticism	**-0.32****	**-0.29****	**-0.24****	**-0.31****	**-0.22****	1.00	0.13	-0.11	0.00	**-0.41****
Open	0.11	0.11	**0.18***	**0.15***	**0.29****	0.13	**1.00**	**-0.33****	**-0.22****	-0.03
Agreeable	0.04	-0.05	-0.07	-0.03	**0.25****	-0.11	**-0.33****	1.00	0.07	0.06
Self-control	-0.12	-0.13	**-0.14***	**-0.14***	**-0.17***	0.00	**-0.22****	0.07	**1.00**	**0.18***
SSI Score	0.10	**0.16***	**0.14***	**0.15***	0.09	**-0.41****	-0.03	0.06	**0.18***	1.00

Note. Correlations marked * are significant at p < .05, those marked ** are significant at p < .004, N=211, (Casewise deletion of missing data).

Table 4.5 reports correlations amongst the 15FQ+, SSI, and GRT2. These results suggest the relationship between integrity and conscientiousness varies at the facet level. Integrity was not significantly positively correlated with self-disciplined, but was with the facets of restrained (r = .23, p<. 004) and conscientious (r = .19, p<. 05). Both of these facets had small significant relationships with integrity, but only that involving restrained met the Bonferroni threshold. Table 4.5 also reports that integrity lacked significant correlations with the fluid ability scale of abstract reasoning (r = .10, p>. 05). However, integrity did have very small significant relationships with the crystallised indicator of verbal (r = .14, p<. 05) and the composite score of general ability (r = .15, p<. 05).

Table 4.5

Correlation Matrix for 15FQ+ primaries, the GRT2, and SSI

	Abstract	Numerical	Verbal	g	SSI	Empathic	Intellectance	Emotional S'	Dominant	Enthusiastic
Abstract	1	.67**	.68**	.87**	.04	.15*	.40**	.11	-.07	.29**
Numerical	.67**	1	.74**	.89**	.06	.07	.43**	.13	-.05	.10
Verbal	.68**	.74**	1	.91**	.04	.17*	.45**	.11	.00	.12
g	.87**	.89**	.91**	1	.05	.15*	.48**	.14*	-.04	.19**
SSI	.05	.06	.04	.05	1	.02	.08	.35**	.07	-.01
Empathic	.15*	.07	.17*	.15*	.02	1	.23**	-.01	-.00	.37**
Intellectance	.41**	.43**	.45**	.48**	.08	.23**	1	.28**	.21**	.22**
Emotional Stability	.11	.14	.11	.14*	.35**	-.01	.28**	1	.11	.11
Dominant	-.06	-.05	.00	-.04	.07	-.00	.21**	.11	1	.15*
Enthusiastic	.28**	.10	.12	.19**	-.01	.36**	.22**	.11	.15*	1
Conscientious	-.05	-.05	-.03	-.05	.10	.00	.04	.03	.14*	-.20**
Socially Bold	.04	.01	.05	.04	.09	.22**	.29**	.26**	.45**	.41**
Tender minded	.01	-.01	.04	.01	.10	.35**	.11	-.06	.04	.06
Suspicious	-.38**	-.28**	-.32**	-.37**	-.23**	-.15*	-.27**	-.22**	.01	-.21**
Abstract	.12	.14*	.20**	.18*	-.12	.05	.24**	-.15*	.15*	.19**
Restrained	-.01	.05	.00	.02	.18**	.18**	.10	.12	-.24**	-.14*
Apprehensive	-.21**	-.16*	-.13	-.18**	-.22**	.07	-.21**	-.55**	-.15*	-.15*
Radical	.12	.11	.16*	.15*	-.12	-.01	.11	-.17*	.09	.16*
Self-sufficient	-.03	.12	.09	.07	-.14*	-.28**	-.03	-.06	-.06	-.35**
Self-disciplined	-.25**	-.25**	-.31**	-.31**	-.02	-.04	-.09	.01	-.01	-.15*
Tense driven	-.23**	-.22**	-.17*	-.23**	-.20**	-.12	-.21**	-.48**	.18**	-.10

Note. Correlations marked * are significant at p < .05, those marked ** are significant at p < .002, N=211, (Casewise deletion of missing data).

Table 4.5 Continued

	Conscientious	Socially Bold	Tender minded	Suspicious	Abstract	Restrained	Apprehensive	Radical	Self-sufficient	Self-disciplined	Tense driven
Abstract	-.04	.04	.00	-.38**	.12	-.00	-.21**	.11	-.03	-.26**	-.23**
Numerical	-.05	.00	-.01	-.28**	.14*	.05	-.16*	.11	.12	-.26**	-.22**
Verbal	-.02	.04	.04	-.32**	.20**	.00	-.13	.16*	.08	-.31**	-.17*
g	-.04	.03	.01	-.37**	.18*	.01	-.18**	.15*	.06	-.31**	-.23**
SSI	.10	.09	.10	-.24**	-.12	.18**	-.22**	-.12	-.14*	-.02	-.20**
Empathic	.00	.22**	.35**	-.15*	.05	.18**	.07	-.01	-.28**	-.04	-.12
Intellectance	.04	.29**	.11	-.27**	.24**	.10	-.21**	.11	-.03	-.09	-.21**
Emotional Stability	.03	.26**	-.06	-.22**	-.15*	.12	-.55**	-.17*	-.06	.01	-.48**
Dominant	.14*	.45**	.04	.01	.15*	-.24**	-.15*	.09	-.06	-.01	.18**
Enthusiastic	-.20**	.41**	.06	-.21**	.19**	-.14*	-.15*	.16*	-.35**	-.15*	-.10
Conscientious	1	.06	.01	.18*	-.16*	.27**	.11	-.26**	.21**	.37**	-.04
Socially Bold	.06	1	.16*	-.06	.20**	-.09	-.35**	.15*	-.24**	-.04	-.05
Tender minded	.01	.16*	1	-.14*	.27**	.09	.11	.09	-.03	-.04	.04
Suspicious	.18*	-.06	-.14*	1	-.14*	-.04	.23**	.05	.15*	.23**	.23**
Abstract	-.16*	.20**	.27**	-.14*	1	-.23**	.07	.43**	.09	-.28**	.13
Restrained	.27**	-.09	.09	-.04	-.23**	1	-.00	-.31**	.05	.20**	-.37**
Apprehensive	.11	-.35**	.11	.23**	.07	-.00	1	.02	.04	.14*	.35**
Radical	-.26**	.15*	.09	.05	.43**	-.31**	.02	1	-.02	-.33**	.19**
Self-sufficient	.21**	-.24**	-.03	.15*	.09	.05	.04	-.02	1	.05	.01
Self-disciplined	.37**	-.04	-.04	.23**	-.28**	.20**	.14*	-.33**	.05	1	.03
Tense driven	-.04	-.05	.04	.23**	.13	-.37**	.35**	.19**	.01	.03	1

Note. Correlations marked * are significant at p < .05, those marked ** are significant at p < .002, N=211, (Casewise deletion of missing data).

The 15FQ+ facets comprising neuroticism are affected by feelings (opposite of emotional stability), suspicious, apprehensive, and tense-driven. Table 4.5 indicates correlations between these facets are all significant, but of varying strengths, ranging from a small $r = .16$ ($p <. 05$) through to a moderate $r = -.40$ ($p <. 002$). Correlation strength between facets and ability indices of fluid, crystallised, and general ability were from weakest to strongest: emotional stability ($r = .16$, $r = .15$, $r = .18$; $ps <. 05$), apprehensive ($r = -.24$, $p <. 002$; $r = -.14$, $p <. 05$; $r = -.21$, $p <. 05$), tense-driven ($r = -.25$, $p <. 002$; $r = -.17$, $p <. 05$; $r = -.24$, $p <. 002$), and suspicious ($r = -.40$, $r = -.34$, $r = -.39$; $ps <. 002$).

Big Five conscientiousness was negatively correlated with crystallised ability ($r = -.14$, $p <. 05$), but no significant relationship was observed between conscientiousness and fluid ability. Furthermore, correlations between significant ability and conscientious facets varied in strength. The facets comprising Big Five conscientiousness are conscientious, restrained, and self-disciplined. Table 4.5 indicates correlations between these facets and ability indices range from a non-significant $r = .01$ through to a significant and moderate $r = -.35$ ($p <. 004$). The only significant correlations involved the facet of self-disciplined and fluid ($r = -.27$, $p <. 004$), crystallised ($r = -.35$, $p <. 004$), and general ability ($r = -.34$, $p <. 004$). The next subsection briefly examines the significance of differences between some of the correlations previously reported.

Section 4.2.2.2: Significance of Correlation Coefficient Differences

This section details a number of calculations to determine the significance of differences among relevant correlation coefficients. These computations were based on Streiger's (1980) recommendations for both large sample sizes and Type I error control.

The first calculation involved the significant correlation coefficients for integrity scores and the Big Five traits conscientiousness and neuroticism detailed in Table 4.4. The relationship between neuroticism and integrity was significantly stronger than the relationship between conscientiousness and integrity $t(208) = -6.87$, $p <.001$.

The second calculation examined significant correlation coefficients for integrity scores and facets of restrained and conscientious. There was no significant difference in strength of correlations for restrained and integrity or conscientious and integrity $t(208) = 0.49$, $p <.617$.

The third calculation examined significant correlation coefficients for neuroticism and the intelligence indices. There was no significant difference in strength of correlations between neuroticism and abstract reasoning or neuroticism and verbal reasoning $t(208) = -1.16$, $p <.246$.

The fourth calculations examined significant correlation coefficients between neuroticism facets and the abstract/fluid intelligence index. There was no significant difference in strength of correlations between abstract reasoning and emotional stability or abstract reasoning and apprehensive $t(208) = 1.21$, $p <.225$. Yet the relationship for abstract reasoning and tense-driven was significantly stronger than that for abstract reasoning and emotional stability $t(208) = -6.67$, $p <.001$. The relationship for abstract reasoning and suspicious was also significantly stronger than that for abstract reasoning and emotional stability $t(208) = -3.01$, $p <.002$.

The fifth calculation examined significant correlation coefficients for neuroticism facets and the verbal/crystallised intelligence index. There was no significant difference in strength for correlations between verbal reasoning and emotional stability or verbal reasoning and apprehensive $t(208) = 0.15$, $p <.874$. Nor was the difference in strength significant for correlations involving verbal reasoning and emotional stability or verbal reasoning and tense-driven $t(208) = -0.29$, $p <.764$. However, the relationship for abstract reasoning and suspicious was significantly stronger than that for abstract reasoning and emotional $t(208) = -2.33$, $p <.02$.

The sixth calculation examined significant correlation coefficients between self-disciplined and intelligence indices. There was no significant difference in correlation strength for self-disciplined and verbal reasoning or self-disciplined and abstract reasoning $t(208) = -1.17$, $p <.204$.

Section 4.2.2.3: Multiple Regressions

The first regression model in Table 4.6 examined the facets of conscientious, restrained, and self-disciplined. This combination was able to account for 6% of the variance in integrity scores. Significant predictors were the facets conscientious ($\beta = .40$) and restrained ($\beta = .55$). There is 95% confidence that the squared multiple correlation is between .01 and .15.

The second model combined the Big Five traits of conscientiousness, neuroticism, and agreeableness. This combination was able to account for 19% of the variance in integrity scores. Significant predictors were neuroticism ($\beta = -1.02$) and conscientiousness ($\beta = .52$). There is 95% confidence that the squared multiple correlation is between .10 and .29.

The third regression model included the four neuroticism facets. This combination was also able to account for 19% of the variance in integrity scores. Significant predictors were the facets emotional stability ($\beta = .80$) and suspicious ($\beta = -.46$). There is 95% confidence that the squared multiple correlation is between .11 and .29.

Table 4.6

β values for Multiple Regression Coefficients of Personality Traits on Integrity

Trait	Integrity β
Conscientious	0.40*
Restrained	0.55*
Self-Disciplined	-0.26
Regression model 1	$F(3,207) = 5.87$
R	0.28
R^2	0.08
Adj. R^2	**0.06**
Conscientiousness	0.52*
Neuroticism	-1.02*
Agreeableness	-0.04
Regression model 2	$F(3,207) = 17.65$
R	0.45
R^2	0.20
Adj. R^2	**0.19**
Emotional Stability	0.80*
Suspicious	-0.46*
Apprehensive	-0.01
Tense-driven	-0.16
Regression model 3	$F(4,206) = 13.39$
R	0.45
R^2	0.20
Adj. R^2	**0.19**
Emotional Stability	0.81*
Suspicious	-0.56*
Conscientious	0.41*
Restrained	0.06
Regression model 4	$F(4,206) = 17.91$
R	0.51
R^2	0.26
Adj. R^2	**0.24**

Note. βs' marked * are significant at $p < .006$

Table 4.6's fourth and final regression model examined significant facet predictors from neuroticism and conscientious. Emotional stability ($\beta = .81$), suspicious ($\beta = -.56$), and conscientious ($\beta = .41$) were able to account for 24% of the variance in integrity scores. There is 95% confidence that the squared multiple correlations are between .16 and .36. The next section presents analyses undertaken to investigate the second research objective of this dissertation.

Section 4.2.3: Adverse Impact Risk and Reduction for those identifying as Māori

This section presents results relevant to the present dissertation's second research objective. More specifically, it investigates whether there is a difference in personality or ability test scores between those identifying as Māori (N = 46) and non-Māori (N = 165). It then explores whether any such differences could be reduced by combining ability tests with other assessments. The results detailed are not expected to resolve issues relating to adverse impact. The results are instead intended to add to the empirical foundations that may allow future assessment and selection practices to more radically reduce adverse impact.

Section 4.2.3.1: Mean Score Differences

Although tentative due to participant group size restrictions, Table 4.7 suggests Cohen's d for those identifying as Māori and broader participant group differences on ability indices were d =.63, $p < .001$ (verbal), d =.47 (numerical), d =.32 (abstract), and d =.54, $p < .001$ (g). Ferguson (2009) suggests that d scores over .41 indicate a practically significant difference in group data. The results in Table 4.7 suggest those identifying as Māori and the broader participant group have significant and practically important group differences on crystallised and overall ability indices (overall index equals combined scores of verbal, numerical, and abstract reasoning scales). However, the score of .32 and accompanying significance test for abstract reasoning suggests that no strong or significant group difference exists between Māori and the broader participant group on this fluid ability index. Table 4.7 also reports an absence of any significant or strong group differences between those identifying as Māori and the broader participant group on integrity test (d = .21) or neuroticism outcomes (d = -.02). Due to sample size restrictions supplementary non-parametric tests were also run. The results of these did not deviate from the preceding findings. In order determine whether differences may be obscured by combining all other ethnicities into a "Non-Māori" group a one-way ANOVA was undertaken. Neuroticism scale outcomes did not differ significantly across any of the ethnic groups differentiated in Table 4.3 (demographics), F (7, 203) = 1.264, p = .270.

Table 4.7

Comparison of Māori and broader participant group Scale Means and Standard Deviations

Scale	Māori Mean	Māori SD	N	Group Mean	Group SD	N	Cohen's *d*
Verbal Reasoning	16.74	5.60	46	20.69	6.84	165	**0.63****
Numerical Reasoning	9.72	5.47	46	12.50	6.17	165	0.47
Abstract Reasoning	12.52	5.94	46	14.47	5.91	165	0.32
g	38.98	14.63	46	47.67	17.01	165	**0.54****
Extroversion	14.78	4.60	46	15.41	3.97	165	0.14
Openness	15.43	3.21	46	16.24	4.15	165	0.21
Agreeableness	-4.23	3.31	46	-4.06	3.65	165	0.04
Conscientiousness	22.85	3.15	46	22.57	4.36	165	-0.07
Neuroticism	5.41	4.62	46	5.29	4.54	165	-0.02
Integrity	38.00	10.35	46	40.30	11.26	165	0.21
SSI Credibility	6.91	2.13	46	7.21	2.17	165	0.13

Note. Differences marked * are significant at p < .05, those marked ** are significant at p < .001

Section 4.2.3.2: Adverse Impact Reduction Calculations

The likely reduction in adverse impact achieved by combining ability, neuroticism, conscientiousness, and integrity scales is based on the group differences reported in Table 4.7 and scale correlations reported in Table 4.4. The potential reduction in adverse impact was calculated via Sackett and Ellingson's (1997) formula:

$$d = \frac{\sum_{i=1}^{k} di}{\sqrt{k + k(k-1)rii}}$$

In this formula, di indicates the d (group difference) value for each predictor included in the composite, k indicates the number of predictors combined to form the composite, and rii indicates the average intercorrelations among the predictors included in the composite.

$$d = \frac{0.54 + 0.21 + -0.07 + -0.02}{\sqrt{4 + 4(4-1)0.24}} = 0.24$$

This calculation suggests that combining ability, neuroticism, conscientiousness, and integrity scales is likely to reduce the group difference between those identifying as Māori and broader participant group scores from d = .54 when ability is used in isolation to d = .24. The next section discusses the implications of these results and others presented in this chapter.

Section 4.3: Discussion (Study 1)

This section discusses the preceding results in light of the research objectives detailed in Chapter 3. The convergence and divergence of findings with previous research is discussed. Speculation as to why these findings may have occurred and a discussion of the consequent implications is also undertaken. The findings are also discussed regarding the resolution of current issues and what they suggest about possible new issues. Theoretical, empirical, and applied implications for the findings are also discussed. This includes implications pertaining to composite design and more general adverse impact and validity calculations. Specific limitations and directions for future research are also discussed.

Section 4.3.1: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability.

The analyses undertaken in Section 4.2.2 investigated whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. Furthermore, they explored variations among any such differences at the facet-level or ability subcomponent.

A variety of international reviews and investigations have reported moderate to strong positive relationships between integrity and conscientiousness (e.g., Murphy, 1993; Ones, 2003; Sackett & Wanek, 1996; Wanek, Sackett, & Ones, 2003). Integrity was also found to positively correlate with conscientiousness within this investigation. However, the relationship was weak and failed to meet the Bonferroni threshold for significance. Facet level relationships between conscientiousness and integrity also failed to clearly replicate previous findings.

Wanek (1995, 1999) argues that the strong relationship between integrity and conscientiousness is due to a strong relationship between integrity and the conscientiousness facet of self-control. The present investigation found integrity positively related to the conscientiousness facets of restrained and conscientious, but *not* self-disciplined. The combination of restrained and conscientious was able to account for 6% of variance in integrity scores. No additional variance was explained by including self-disciplined in the regression model. Wanek (1999) suggested that integrity tests tend to place a lot of emphasis on self-control while Big Five measures of conscientiousness generally place greater emphasis on orderliness, perseverance, and conformity. He hypothesised that this different focus represented something akin to a failure of either test to adequately cover the full domain of conscientious behaviour. Wanek felt this accounted for the strong, but less than perfect relationship between integrity and conscientiousness. Sackett and Wanek (1996) hypothesised that it was this differential emphasis on self-control that is responsible for the incremental validity integrity tests contribute to conscientiousness alone. Yet the conscientiousness facet of self-disciplined was the only facet of this trait *without* a significant relationship with integrity test scores (self-disciplined is the 15FQ+ synonym for self-control).

It could be argued that 15FQ+ facets of restrained or conscientious are better representations of Wanek's (1999) conceptualisation of self-control than self-disciplined is. As noted in section 4.1.2, the 15FQ+ conscientious facet correlated most strongly with the 16PF- 5 trait of self-disciplined, while the 15FQ+ self-disciplined facet correlates most strongly with the 16PF - 5 trait of conscientiousness. This was attributed to a reversal of these factors in the latest edition of the 16PF. Yet the reversal of these factors could be taken as support for the conscientious facet being a better approximation of Wanek's self-control than the self-disciplined facet. Support for this explanation would have come from a stronger relationship between integrity and the conscientious facet than integrity and the restrained facet, but no such relationship was observed.

While this investigation did not find the expected relationship between self-control and integrity, its findings do still support the idea that the key to understanding the relationship between integrity and Big Five conscientiousness lies in its facets (Sackett & Wanek, 1996). Furthermore, while a lack of support for the relationship between integrity and self-control may affect how we explain the relationship between integrity and conscientiousness, it does not affect what is known in the philosophy of science as the *theoretical core* of any such explanation (Lakatos, 1968). In other words, the findings of this investigation do not affect the idea that there *is* a relationship between conscientiousness and integrity. They instead affect the more peripheral components of the theory that try to explain how this relationship operates. They simply suggest that the explanations for the relationship between conscientiousness and integrity that focus on self-control may need to be replicated and better understood before any serious conclusions are drawn.

Explanations for employee delinquency not relying on a lack of self-control are also well substantiated. Greenberg (1998) has written extensively on what he calls the cognitive geometry of employee theft. According to Greenberg employees who steal from the workplace use a number of tactics to dissuade both others and themselves from interpreting their “taking” behaviour as “theft.” Greenberg conceptualises both the cognitive/behavioural and impression management sequence of employee theft as follows: An employee makes the decision to take property, this decision is then justified, the property is then taken, the interpretation of this act is then managed and the act is labelled as “taking” rather than “theft.” This theory of delinquency suggests that employee theft is a consequence of a cognitive juggling act that justifies the act of taking, rather than knowing that such an act is wrong, but not having the self-control (ego-strength) to resist one’s impulse to attain immediate gratification. Both of these explanations of counterproductive behaviour are likely to explain various episodes of workplace delinquency. However, the lack of significant relationship between self-control and integrity suggests that Greenberg’s cognitive explanation may predominate. It also suggests that people with high integrity may not perceive situations where dishonest behaviour is possible as opportunities for counterproductive workplace behaviour. As a result there is less need to use self-control to avoid engaging in the dishonest behaviour.

Examining conscientiousness’s facet traits was also important for investigating the homogeneity of the conscientiousness construct. It has been argued that Big Five conscientiousness is comprised of traits of varying congruence regarding behavioural tendencies (Tett et al., 1991). The existence of such incongruence has been found to obscure important linkages among job performance and facet traits, and to reduce the predictive accuracy of personality traits for job performance (Tett, Steele, & Beauregard, 2003). On this basis an exclusive focus upon broad-level traits may reduce the incremental validity gained via the addition of conscientiousness and integrity to cognitive ability. The simplest way to test this incongruence was to explore differences in the positive or negative loading of conscientiousness facets upon integrity. The findings suggested that the relationship among conscientiousness facets and integrity loaded consistently.

Previous researchers have explained the relationship between conscientiousness and integrity via a superordinate personality or developmental factor responsible for these propensities (e.g., Ones et al., 2005a; Wanek, 1999). One superordinate personality factor previously discussed is Digman's (1997) Alpha factor. Alpha factor maturation is hypothesised to influence the development of impulse control and conscience. It is also thought to positively influence reductions in hostility, aggression, and neurotic defense. If normal childhood maturation occurs a child develops a superego and learns to restrain or redirect id impulses and to discharge aggression in socially approved ways. Yet if socialisation is hindered for any reason, Alpha factor development is retarded, and neuroses, superego deficiency, and/or excessive aggressiveness will result (Digman, 1997). This suggests that predispositions towards integrity, conscientiousness, and neuroticism travel a similar if not identical developmental trajectory. Digman's broad account of Alpha structure and development also includes Big Five's agreeableness factor (Rushton & Irwing, 2008, 2009). The results did not find a significant relationship between integrity and agreeableness. This suggests that conscientiousness and neuroticism are more important influences on the tendency to engage in counterproductive behaviours.

To further test the relationship between Digman's (1997) Alpha factor and integrity multiple regressions were undertaken with Big Five conscientiousness, neuroticism, and agreeableness as predictors. These regressions found conscientiousness and neuroticism able to explain 19% of variance in integrity test scores. Yet the addition of conscientiousness and agreeableness was found to account for *no* additional variance to that of neuroticism facets alone. This fails to provide support for Digman's shared developmental trajectory, or account for previously observed relationships among conscientiousness and integrity via the Alpha factor. Instead it suggests that previously observed relationships between integrity and conscientiousness may have primarily occurred as a function of the moderating effect of neuroticism traits on both constructs. This is further supported by the finding that the facets of emotional stability and suspicious were able to explain 19% of variance in integrity scores.

In line with these regression outcomes, the results of this investigation replicated international findings suggesting a negative relationship between neuroticism and integrity (e.g., Gough, 1972; Hogan & Hogan, 1989; Ones, Schmidt, & Viswesvaran, 1993; Woolley & Hakstian, 1992). This is consistent with the relationship between neuroticism and counterproductive behaviour discussed in section 2.3.3. Furthermore, a regression equation combining the facets of emotional stability, suspicious, conscientious, and restrained found significant predictive relationships for the former three only. This regression was able to account for 24% of variance in integrity test scores, which is the greatest variance accounted for by any combination of predictors examined. This suggests that facets from Big Five neuroticism and conscientiousness provide a better prediction of integrity variance than any other combination of traits. This again supports the importance of examining facet level relationships over and above those of at the Big Five level.

Results also examined relationships between integrity and ability and again supported the importance of examining facet level relationships. The findings of this investigation suggested that integrity scores were not related to fluid or general ability indices, but did have a very small negative relationship with the crystallised index of verbal reasoning. However, the very small size of this relationship indicates that respondents' integrity test scores are not likely to systematically alter to any practical extent as a function of their ability. On this basis the risk of CWBs does not appear greatly affected by educational attainment or innate intellectual capability. One explanation for this finding is that the psychological defence mechanism of rationalisation is not related to intelligence. Defence mechanisms such as *rationalisation* are a major, relatively unconscious means of managing conflict and affect (A. Freud, 1937). Rationalisation is thought to manage potential internal conflict when engaging in counterproductive workplace behaviours by allowing protagonists to normalise and justify their behaviour (Greenberg, 1998). The lack of relationship between rationalisation and intelligence explanation is supported from more general findings in this area. Cramer (1999) found that intelligence and defense mechanisms both relate to ego development, but are themselves unrelated. This suggests that even the most intelligent of people answering integrity tests are unlikely to realise how deviant their behaviour is. In other words, even those who are "smart" enough to readily discern the nature of test questions may still fail to recognise how aberrant their answers may appear.

The *false consensus effect* provides another potentially inclusive explanation for this lack of relationship between integrity and intelligence. According to Murphy (1993) integrity tests are based upon the idea that less honest individuals are likely to estimate a higher incidence of dishonesty in the general population. This tendency would support the idea that integrity is not related to intelligence because the belief in a greater general incidence of dishonest behaviour, leads dishonest people of all levels of intelligence to mistakenly believe their answers are consistent with the norm. In other words, a more intelligent deviant might wish to distort their answers to appear "honest," but they underestimate the difference between their true answers and those of "real" honest people. This appears consistent with the false consensus effect. According to the false consensus effect people are biased when it comes perceiving the normalcy of their own behaviours, attitudes, and responses (Alicke & Largo, 1995; Ross, Greene, & House, 1977). Research into the false consensus effect has shown that those answering questions, such as those contained with integrity tests, believe their own responses to be relatively common and thus relatively unrevealing. On the other hand, respondents with CWB propensities are likely to consider responses that differ from theirs to be relatively uncommon and thus revealing (Ross et al., 1977). The existence of the false consensus effect is well supported empirically (de la Haye, 2000; Marks & Miller, 1987). The theory that the false consensus effect might increase the perceptions of normalcy for deviant behaviour is also specifically supported. For example, cannabis and amphetamine users make significantly higher estimates of the prevalence of drug use among students than do nonusers (Wolfson, 2000).

The finding that no practically significant relationship was evident for both fluid or crystallised ability and integrity also supports the inclusion of both predictors within the composite developed in Study 2 and make a contribution to the literature. Fluid ability assesses largely innate capability. Crystallised ability loads heavily upon acculturation and education. This is an important distinction to make in applied settings, yet one apparently overlooked in previous research into relationships between integrity and ability. Being able to distinguish between Gf and Gc facilitates predictions of training success versus immediate task performance capability. Previous research examining relationships between personality traits and Gf/Gc found divergent relationships (Moutafi et al., 2003; Wood & Englert, 2009). These findings suggested that relationships amongst personality traits and ability components may differ according to cognitive and situational requirements. While more work is required before strong claims can be made, the finding that integrity differed in its relationships between Gc and Gf further supports the importance of this distinction.

The present investigation also found small to moderate significant negative relationships between neuroticism and crystallised, fluid, and *g* indices. This was consistent with previous findings (Ackerman & Heggestad, 1997; Austin et al., 2002; Hembree, 1988; Kyllonen, 1997; Seipp, 1991). The previously discussed neural efficiency hypothesis provides one potential explanation for these findings. This suggests that respondents experiencing relatively high levels of anxiety engage in substantially more task irrelevant processing (i.e., worry) (Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Zeidner & Matthews, 2000). Another way to think of this is that respondents with more neurotic tendencies make less efficient and effective use of their cognitive resources when undertaking ability assessments.

The strength of relationships between neuroticism and ability varied from small to moderate across facets. Suspicious and affected by feelings were the facets most strongly associated with differences in ability index magnitudes and the only ones able to account for unique variance in ability scores. The ability indices small to moderate relationship with affected by feelings and the small and marginally significant relationship with apprehensive are broadly consistent with previous research suggesting that depressive tendencies are associated with impaired reading comprehension and problem solving (Hartlage, Alloy, Vasquez, & Dykman, 1993). The finding that the ability indices were most strongly related to the facet of suspicious is less consistent with expectations. The suspicious scale is comprised of questions such as “everybody has their price” or “most people only really care about themselves”. These are questions that primarily assess cynical attitudes rather than the emotive sentiments of most other neuroticism facets. This brings into question the neural efficiency hypothesis as an explanation for the negative finding between intelligence indices and Big

Five neuroticism. On the basis of this explanation the most emotive facets would be expected to have the strongest relationships with ability indices, but relationships with the least emotive facet were significantly stronger than those of facets more consistent with this explanation. A replication of this finding in Study 2 may have important implications for the neural efficiency hypothesis as an explanation for the negative relationship between ability indices and neuroticism.

Relationships between conscientiousness and ability indices were also observed. The crystallised ability index was negatively related to conscientiousness scores, but no relationship was observed with the fluid ability index. This finding is consistent with previous research examining the relationship between conscientiousness and ability indices within a New Zealand participant group (Wood & Englert, 2009). However, it is inconsistent with previous U.K. based findings that suggested conscientiousness is negatively correlated with fluid, but not crystallised ability (Moutafi, et al., 2003, 2004). The difference in observed relationships across U.K. and New Zealand investigations has been explained as a function of differences in the educational/achievement level of their respective participant groups (Wood & Englert, 2009). According to Wood and Englert the negative relationship between conscientiousness and fluid ability found by Moutafi et al. was a reflection of their having used a highly educated "achievement striving" participant group. It was thought that any ability differences in such a similarly educated participant group would reflect differences in learning ability rather than knowledge. The current study's use of archival data for a general reasoning test battery makes this explanation equally applicable now.

The finding that crystallised ability and conscientiousness are negatively related provides further support for the intelligence compensation theory (ICT) (Wood & Englert, 2009). This theory suggests that behaviours associated with conscientiousness will be reinforced for individuals of lesser fluid or crystallised intelligence relative to their cohort, and on this basis act as a compensating mechanism for those with lesser ability. A replication of this finding in Study 2 will provide additional support. However, this theory will be brought into doubt if Study 2 finds that scores on conscientiousness have a relationship with how long it takes people to complete ability assessments. If more conscientious people tend to take longer to complete ability scales then the negative relationship between conscientiousness and crystallised ability may reflect an artefact of test-taking style rather than a relationship between conscientiousness and actual ability. If the test-taking style explanation of the negative relationship between conscientiousness and crystallised ability is supported, an explanation for why this test-taking style doesn't negatively impact upon fluid ability scores will be required.

Relationships between conscientiousness and ability indices were also found to differ at the facet level. Self-disciplined facet scores were more strongly associated with the magnitude of ability scores than other conscientiousness facets. Small to moderate relationships were observed between self-disciplined and both fluid and crystallised indices. No difference in the magnitude of ability indices was associated with the magnitude of the conscientious facet. The finding that relationships between conscientiousness and intelligence indices vary at the facet-level is also consistent with the research discussed concerning the preceding hypothesis (Moutafi et al., 2004; Wood & Englert, 2009). Yet the lack of a significant relationship with the facet of conscientious is not.

The finding that relationships differ at the facet level has been a consistent and important theme to emerge regarding the first research objective. The next section will discuss the outcomes in reference to the second research objective of this dissertation.

Section 4.3.2: Adverse Impact Risk and Reduction for those identifying as Māori

Findings indicated those identifying as Māori scored substantially worse than those from the broader respondent group on the ability indices of Gc and *g*. While differences were observed on all indices, such differences were only significant for crystallised and general ability. This suggests that there are noteworthy differences between those identifying as Māori and the broader respondent group on knowledge-based components of intelligence, but less so for scales assessing novel problem-solving potential (fluid index). The size of this difference suggested that if those identifying as Māori take part in selection processes depending upon ability test outcomes they may experience adverse impact.

The finding that those identifying as Māori scored relatively poorly on crystallised ability appears consistent with a previous finding in this area (Guenole et al., 2003). Guenole et al. found that significant differences between Māori and NZ Europeans/Pakeha on ability scales were very large for the numerical business assessment and large for the verbal reasoning tests, but not significant for the general numerical test. This appears to be consistent with this investigation's finding that those identifying as Māori did relatively better on a fluid intelligence index, for general numerical reasoning tests often load heavily on fluid intelligence (Ackerman, 1997).

Those identifying as Māori were found to have no significant or practical differences with the broader participant group for the integrity assessment. The current investigation represents the first research undertaken into integrity test differences between those identifying as Māori and a broader New Zealand participant group. The finding that integrity test outcomes do not differ significantly between these groups is consistent with U.S.A. based research findings. Such research has consistently indicated no significant differences in integrity outcomes between majority and minority group members (Ones & Viswesvaran, 1998a; Sackett et al., 1989; Strand & Strand, 1986; Terris, 1983, 1985; Viswesvaran & Ones, 1997).

No significant differences were found for the other personality traits reported either. The lack of significant difference for neuroticism scores suggests that those identifying as Māori are no more likely than the broader participant group to be suspicious of other people's motives, experience social anxiety or self-doubt, display impatience, or be emotionally volatile. This is inconsistent with previous work examining ethnic differences on personality outcomes within New Zealand. Packman et al. (2005) found significant effect size differences between those identifying as Māori and New Zealand Europeans on the Big Five trait of neuroticism (d = .30, p < .05). One difference between the current investigation and that of Packman et al. is the strength of d threshold employed. While this investigation follows Ferguson's (2009) relatively restrictive .41 minimum threshold for relevant group differences, Packman et al. followed Cohen's (1988) .2 threshold for noteworthy, albeit small differences. Yet the d value of -.026 reported in this study fails to meet even Cohen's minimum threshold, which rules out this as an explanation for inconsistent findings.

Another possible explanation for the inconsistent finding with Packman et al. (2005) is that this investigation compared personality test outcomes between those identifying as Māori and the broader participant group, rather than against New Zealand Europeans. This creates the potential for differences between those identifying as Māori and NZ European to be obscured by the moderating impact of other ethnicities included within the "broader participant group." However, this was not supported by the ANOVA undertaken to test this possibility.

The finding that Māori differ from non-Māori on ability results suggests that Māori are likely to experience adverse impact in selection systems involving ability assessments. The strength of relationships between the scales examined and the lack of significant differences on personality assessments also suggests that such adverse impact may be reduced through combining ability assessments with those of integrity, neuroticism, and conscientiousness.

The size of the difference between those identifying as Māori and the broader participant group on ability scores suggests in ability-based selection systems Māori are likely to be employed at a ratio of less than 1:5 when 1% of non-Māori applicants are successful. In situations where 50% of non-Māori applicants are successful, only 30% or less Māori applicants are likely to be hired (Sackett & Ellingson, 1997). One way in which to evaluate these differences is by looking to international requirements and heuristics and the Treaty.

A common way of considering the impact of group differences on the hiring rates of minority group members is by applying what is known as the "four-fifths" rule of thumb. This is the heuristic used by the North American Equal Employment Opportunity Commission to determine *prima facie* cases of adverse impact (*Uniform Guidelines,* 1978). According to this heuristic the percentage of minority job applicants must be selected at a minimum of 80% (four-fifths) of the percentage of majority job applicants selected. Selection decisions that are less than this four-fifths percentage are considered discriminatory unless based upon proven job-specific performance predictors. Although New Zealand corporations are not bound by the four-fifths rule, in the interests of ethical practice and the achievement of goals pertaining to workplace diversity, it can be a useful heuristic to adopt. Using the hiring ratios conversion information presented in Sackett and Ellingson (1997) as a guide, 90% of non-Māori applicants would have to be accepted before those identifying as Māori would be selected in-line with the four-fifths requirement.

Through combining integrity, conscientiousness, and neuroticism with ability results it is possible to substantially increase the hiring ratios of Māori applicants. The calculation undertaken suggests the magnitude of group differences between those identifying as Māori and the broader participant group can be reduced from .547 when ability is used in isolation, to .249 when ability is used in conjunction with integrity, neuroticism, and conscientiousness. With this reduced difference those identifying as Māori would now be hired at a ratio of 3:5 in situations where 1% of non-Māori applicants are successful. In situations in which 50% of non-Māori applicants are successful, 42% of those identifying as Māori applicants are likely to be hired (Sackett & Ellingson, 1997). This represents an increase in hiring ratios for those identifying as Māori of 200% and 40% respectively. When translated into a "four-fifths" context, those identifying as Māori will now be hired at the rate of 80% of non-Māori applicants when only 50% of non-Māori applicants are accepted.

The preceding discussion is not intended to resolve issues relating to adverse impact. However, it is hoped that it will build the theoretical and empirical foundation that will allow future assessment and selection practices to more radically reduce adverse impact. The next subsection examines specific limitations of Study 1. More general limitations are discussed in section 6.2.

Section 4.3.3: Specific Limitations

The measures used in Study 1 present a number of limitations specific to this study. Firstly, while the GRT2 is an adequate operationalisation of intelligence measured by tests, all its dimensions are designed to load heavily on fluid ability - including its verbal reasoning scale (Budd, 1993). This means its verbal reasoning component may not actually be a particularly good index of crystallised intelligence.

The GRT2 was also designed for the general population. This means it is unlikely to finely discriminate among higher-level scorers. The narrow scope of this design may be further exacerbated through the use of a participant group predominantly comprised of pre-management applicants. The reason this presents a limitation is that variations in the strength of associations among factors such as personality traits and ability are likely across subgroups of different ability level (Austin et al., 2002) and have been demonstrated elsewhere (Wood & Englert, 2009).

Using a participant group of job applicants is beneficial for examining relationships in a "real-life" context (Borman & Motowidlo, 1997; Guba & Lincoln, 1994; Landy, Shankster & Kohler, 1994; Ones & Viswesvaran, 2001a; Wanek, 1999). Yet using actual job applicants raises the possibility that motivated distortion of the results may have affected outcomes. Considerable evidence argues against this likelihood (Barrick & Mount, 1996; Christian, Goffin, Johnson, & Rothstein, 1994; J. Hogan, Barrett, & R. Hogan, 2007; Ones et al., 1993b; Smith & Ellington, 2002). Nevertheless, undetected distortion must still remain a concern and its potential influence on findings is worthy of consideration (Kline, 1995, 2000; Lao, 2001; Mischel, 1968).

Section 4.3.4: Future directions

It was suggested that Greenberg's (1998) rationalisation explanation of CWBs was more consistent with observed relationships than previously popular self-control/impulsive control explanations (Murphy, 1991). If accurate, this may have important implications for the development and implementation of theft reduction strategies. For example, strategies that involve reducing ambiguity over the appropriateness of counterproductive behaviour may be more effective than simply attempting to limit employee opportunities for CWBs. In order to test the generalisability of self-control's lack of correlation with integrity future research may benefit from examining a participant group with histories of dishonesty.

Section 4.3.5: Summary

This study set out to investigate whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated within New Zealand. The relationships observed among, integrity, conscientiousness, neuroticism, and ability were largely consistent with previous findings (e.g., Ones, 1993). Conscientiousness and neuroticism facets were able to account for more variance in integrity scores than the broader scales themselves, and relationships between conscientiousness and ability was found to differ according to the ability indices examined. This reconfirms and further supports the importance of focusing on facet-level traits when exploring relationships among personality traits and other individual difference factors.

The results also suggested that the constructs of integrity and conscientiousness are theoretically and empirically distinct. The strength of intercorrelations among traits measured appeared unlikely to greatly affect their utility if combined within a composite of predictors. This supports Study 2's development of a composite of these predictors. Furthermore, the information presented can facilitate the computation of incremental validity estimates based on combining personality traits, intelligence, and integrity (Ones & Viswesvaran, 2001a, 2001b, 2005c).

This chapter also attempted to investigate potential differences in personality and ability test scores between those identifying as Māori and non- Māori. Differences were observed on ability assessments, but not for integrity or other personality traits. The differences in ability suggested that adverse impact may occur for Māori. However, calculations into the potential to reduce this adverse impact suggested hiring ratios can be increased by combining ability tests with other assessments (Sackett & Ellingson, 1997). The next study in this dissertation attempts to replicate the findings of this study with different scales and participants. It also investigates the potential mechanisms underlying observed relationships between predictors by examining the place of response time in assessment outcomes.

Chapter 5: Study 2

This chapter reports on the creation of a new composite assessment of important general predictors of job performance. The composite is comprised of scales of conscientiousness, neuroticism, integrity, and fluid and crystallise ability. The rationale for the creation of this composite is two-fold. Firstly, it allows for unfettered access to item to scale relationships and the testing of scale redundancy/structure, which is important for construct clarity and understanding any relationships observed. Secondly, it has the potential to contribute in a novel way to existing assessments in two respects. Firstly, it attempts to reduce the potential for ability assessment use to result in adverse impact for those identifying as Māori. Secondly, it attempts to gather response time data clarifying optimal administration and interpretation parameters. This chapter also investigates the research objectives detailed in Chapter 3:

The first two research objectives concerning predictor interrelationships and Māori differences in average score were previously investigated in Chapter 4. Study 2 will look to replicate and briefly report on these findings. The third research objective to investigate whether the length of time taken to answer ability assessments is related to respondent ability or item difficulty and explore whether differences in conscientiousness or neuroticism scores meant respondents take longer to answer ability assessments will be investigated and discussed for the first time in the present study. Section 5.1 details the research design, procedure, scale development, and analytic choices made. Section 5.2 then outlines the results of this investigation. Section 5.3 then discusses these results concerning scale characteristics, previous findings, and wider contributions to the field in general.

Section 5.1: Method

The current investigation employed a correlational, non-randomised, and within-subjects design. Data collected was cross-sectional. The earlier sections and/or subsections of this chapter focus on assessment development. The latter sections and/or subsections focus on investigating the previously outlined research objectives.

The first step in test construction involved specifying and writing items and then administering them to a participant group. A number of analyses were then performed on the derived data to facilitate item quality and the creation of potentially valid and reliable scales. The choice and rationale for the specific statistical procedures used are detailed in the “analysis” subsection. Test construction occurred within both classical test theory (CTT) and item response theory (IRT) statistical frameworks.

Section 5.1.1: Participants (Study 2)

Three hundred and seventeen respondents participated in Study 2. Generic email invitations were sent out to potential participants via I/O Net, the Massey University “Psych Grad” list, and the researcher’s Facebook account. These invitations contained details of the pilot test content, logistics for completion, and purpose of the project. They also contained a link that potential participants could use to anonymously access the pilot assessment via the internet. A link to the pilot assessment was also posted on OPRA Group’s website. The link and accompanying information were provided in OPRA’s section on candidate preparation for completing psychometrics. Data were collected from 16 June 2010 through till 15 September 2010. Demographic information for these participants is detailed in Table 5.1.

Demographic taxonomies in Table 5.1 are consistent with those collected by Statistics New Zealand’s 2006 census. A comparison of the 2006 census and demographic distributions in Table 5.1 suggest a number of discrepancies. These include an underrepresentation of both males and Māori, which made up 48% and 14.6% of the NZ population in 2006. The largest difference between this investigation’s participant group demographics and those outlined in the 2006 census relates to educational levels. The 2006 census reports 40% of New Zealanders with a post-school qualification. Table 5.1 suggests 79.5% of this investigation's participant group had a post-school qualification. Five percent of the NZ population were reported to have a higher degree in 2006. Table 5.1 suggests 26.2% of the participant group had a higher degree.

Table 5.1

Demographic Statistics, Pilot Testing Investigation Participant Group

N = 317	Frequency	%
Gender		
Male	118	37.2
Female	194	61.2
Missing	5	1.6
Ethnicity		
African	2	0.6
Asian	16	5.0
Australian	17	5.4
European	29	9.1
Latin American	2	0.6
Māori	22	6.9
Middle Eastern	2	0.6
NZ European/Pakeha	196	61.8
Pacific Peoples	4	1.3
South African European	4	1.3
Other	18	5.7
Missing	6	1.9
First Language		
English	285	89.9
Other	26	8.2
Missing	6	1.9
Age		
16-20	6	1.9
21-25	72	22.7
26-30	65	20.5
31-35	65	20.5
36-40	39	12.3
41-45	20	6.3
46-50	14	4.4
51-55	14	4.4
56-60	12	3.8
60+	5	1.6
Missing		
Education		
Less than Secondary School	3	0.9
Secondary School	57	18.0
Trade/industry Qualifications	41	12.9
Vocational Qualification	17	5.4
Bachelor Degree	98	30.9
Higher Degree	83	26.2
Other	13	4.1
Missing	5	1.6

Note. No additional information is available to clarify respondent selections of "other."

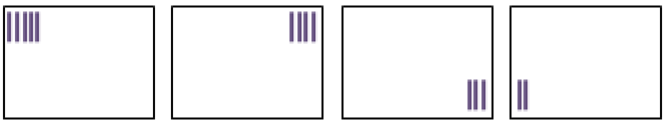
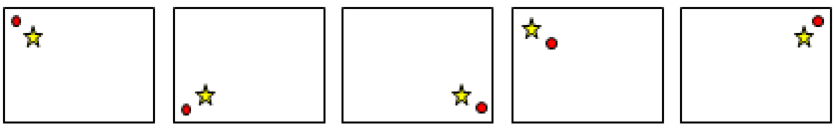
Section 5.1.2: Measures (Study 2)

Item pool establishment: This section details the design of pilot items intended to measure conscientiousness, neuroticism, integrity, and fluid and crystallised intelligence. Although not a focus of this study, distortion/dissimulation items intended to identify potential respondent manipulation were also written and piloted. These dissimulation items are administration items not intended to measure personality trait characteristics. The inclusion of a dissimulation scale is in line with British Psychological Association requirements and test user concerns. Further information to that contained within this section on the rationale for item content is available on request. The primary rationale for designing new items was to ensure item-to-scale knowledge and to provide an opportunity for data-driven structure discovery incorporating full domain coverage.

Items designed for inclusion took into account constraints on intrinsic item properties and item features (Stocking & Swanson, 1993). Issues relating to response option location were minimised through the consistent placement and use of a five-point Likert response option across all items and scales. Wherever possible, items limited the reuse of words, pictures, phrases, or other content across multiple items and distractors. Item design also avoided the use of common stimulus material. Table 5.2 details example items for each of the scales created. Subsequent subsections provide information on the parameters used to guide the design of items for each of these scales. A full list of all items piloted is available on request. Items piloted and retained are contained within Appendix A.

Table 5.2

Example Scale Items

Scale	Example Items
Conscientiousness (50 items)	• I am a well organised person • I like to take life as it comes rather than plan too far ahead* • I am a person of a lot of self-discipline
Neuroticism (50 items)	• I often worry about things • I sometimes wonder if people really like me • I sometimes lose control of my emotions and feelings
Integrity (40 items)	• It is okay to cheat someone who has cheated you first • I would keep the money if I were given too much change in a store • Most people would steal something if they knew they would not get caught
Dissimulation (20 items)	• I always do the right thing • I have never broken a promise • I never judge people before I know them
Crystallised Intelligence (30 items)	• The meaning of prudent is closest to? 1) Guarded 2) Shy 3) Educated 4) Reserved 5) Sensible • Which of the following is the odd one out? 1) Elephant 2) Whale 3) Tiger 4) Dolphin 5) Shark
Fluid Intelligence (30 items)	 What comes next?  Which is the odd one out? 

Note. *Reverse scored item. See Appendix A for retained items. All items are available on request.

Section 5.1.2.1: Personality Item Construction

Personality items were written on the basis of two considerations. The first concerned construct domains documented in the literature review chapter. An examination of extant assessments of relevant constructs provided the second stimulus for item design. This involved examination of 11 assessment test booklets and technical manuals. Of the assessments examined, acceptable reliability and validity formed criteria of weight for inspiration. Item design for the personality predictors also involved a comprehensive examination of the *International Personality Item Pool* (IPIP) http://ipip.ori.org/newMultipleconstructs.htm, which is a "collaboratory" for the development of sophisticated assessments of personality and other individual differences (Goldberg et al., 2006). As well as IPIP Big-Five Factor markers constructs and items were examined from:

- **NEO-PI-R** domains of neuroticism and conscientiousness. Neuroticism facets/scales included: anxiety, anger, depression, self-consciousness, immoderation, and vulnerability. Conscientiousness facets/scales included: self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, and cautiousness.
- **16PF** domains of anxiety and self-control. Anxiety facets/scales included: emotional stability, vigilance, apprehension, and tension. Self-control facets/scales included: liveliness, rule-consciousness, abstractness, and perfectionism.
- **15FQ+** domains of anxiety and control. Anxiety facets/scales included: affected by feelings, suspicious, self-doubting, and tense-driven. Control facets/scales included: conscientiousness, restrained, and self-disciplined.
- **California Personality Inventory (CPI)** domains of well-being and norm-favoring. Well-being facets/scales included: sociability, socialisation, good impression, femininity, and amicability. Norm-favoring facets/scales included: self-control, responsibility, achievement via conformance, flexibility, and law enforcement orientation.
- **Tellegen's Multidimensional Personality Questionnaire (MPQ)** constructs examined included: achievement, stress reaction, aggression, alienation, and control.

- **Myers-Briggs Type Indicator (MBTI)** constructs of extroversion – introversion, sensing – intuiting, thinking – feeling, and judgement – perception.
- **Jung-Type Indicator (JTI)** constructs of extroversion – introversion, sensing – intuiting, thinking – feeling, and judgement – perception.
- **Occupational Personality Profile (OPP)** domains of conscientiousness and anxiety. Facets/scales examined included: detail-conscious, assertive, cynical, phlegmatic, contesting, and pessimistic.
- **Hogan Personality Inventory (HPI)** domains of prudence and adjustment. Due to conceptual similarities the constructs of reliability and stress-tolerance were also examined.
- **Stanton Survey of Integrity (SSI)** domains of work related theft, theft unrelated to work, and company policy violations.
- **London House Personnel Selection Inventory (PSI)** domains of honesty, tenure, and drug avoidance.
- **The Reid Report** domains of integrity attitude, social behaviour, substance use, and work background.

One hundred items were written to assess conscientiousness, 100 to assess neuroticism, 90 for integrity, and 30 for dissimulation. Kline (2000) suggests twice as many items be piloted as are expected to be eventually retained, but that item numbers should not be so great as to result in boredom or fatigue. He further recommends that an hour is the maximum length of time adult respondents to be expected to answer assessments. The initial 100, 100, 90, and 20 items written were therefore reduced to 50, 50, 40, and 20 items respectively. This reduction was also based upon findings that reliability and validity for personality are achievable with four or fewer items per scale (Barrett & Paltiel, 1996; Donnellan, Oswald, Baird, & Lucas, 2006). The refinement was based upon apparent redundancy of behavioural measurement and consultation with subject matter experts within the psychometrics industry. Both subject matter experts have been directly involved in the development of psychometrics and have publications in the area of psychometric design and/or use.

Conscientiousness items were written to encompass behaviour related to a variety of facets. These facets include: *competence* (e.g., I regularly find myself in situations for which I am not fully prepared), *order* (e.g., I prefer to follow a routine), *detail-orientation* (e.g., I notice when things are out of place), *self-disciplined* (e.g., I am a person with a lot of self-discipline), *achievement-striving* (e.g., I work hard to achieve the goals I have set myself), *dependability* (e.g., If I start something, I just about always finish it), and *deliberation* (e.g., I normally think about all my choices before deciding what to do).

Neuroticism items were also written to encompass behaviour related to a variety of facets. These facets were: *anxiety* (e.g., I often feel worried about how things will go in the future), *affected by feelings* (e.g., people have told me I am moody), *anger/hostility* (e.g., it takes a lot to make me angry), *tense-driven* (e.g., I am annoyed when things take longer than expected), *depression* (e.g., sometimes I just feel like staying in bed all day), *self-consciousness* (e.g., I sometimes wonder if people really like me), *self-doubting* (e.g., I sometimes doubt myself), *vulnerability* (I sometimes feel overwhelmed when I am under stress), *suspicious* (e.g., most people can be trusted to keep their promises), and *impulsiveness* (e.g., I find it difficult to resist temptation).

Integrity item writing attempted to facilitate respondents' use of Greenberg's (1998) psychological tactics for CWBs (i.e., minimisation, externalisation, normalisation, and superordination). Integrity items were written to encompass *poor citizenship behaviours* (e.g., I sometimes work more slowly than I could), *active dishonesty* (e.g., taking something from work without permission is not always "stealing"), *enabling attitudes* (e.g., some people have good reasons for stealing), and *company policy violations* (e.g., it is okay to bend company rules when necessary).

The dissimulation items were designed to identify someone claiming unlikely virtues or denying overwhelmingly common faults approaches (Kuncel & Borneman, 2007). An example of the former is "I like everyone I meet" and an example of the latter is "I have never broken a promise".

Section 5.1.2.2: Cognitive Ability Item Specification and Writing

Based upon considerations outlined in the literature review chapter 50 items were written to assess fluid and crystallised intelligence each. Items were then reduced to 30 each scale for piloting. As with the previously detailed personality items, the refinement of the cognitive ability items was to reduce test length and redundancy and occurred in consultation with subject matter experts.

As with the design of personality items, ability item writing was inspired by construct domain and extant assessments. Carroll's (1993) survey of factor-analytic studies into human cognitive ability informed the assessments chosen for review. The general and graduate reasoning tests were examined in addition to those recommended by Carroll's research:

- **Raven's Progressive Matrices**
- **Culture Fair Intelligence Test**
- **Woodcock-Johnson Psycho-Educational Battery**
- **USES General Aptitude Test Battery (GATB)**
- **Kaufman Assessment Battery for Children (K – ABC)**
- **Wechsler Adult Intelligence Scale – Revised (WAIS-R)**
- **Armed Services Vocational Aptitude Battery (ASVAB)**
- **General Reasoning Test Battery (GRT2)**
- **Graduate Reasoning Test Battery (GRT1)**

Based upon this examination and the literature review, fluid intelligence items were written to incorporate items tapping into visualisation, induction, and speed of reasoning. More specifically, items were written to assess the perceptual and cognitive ability enabling a person to deal with spatial relations and the orientation of objects in space. Further assessed was the ability to organise collections of items by finding a rule or set of common characteristics accounting for similarities and/or differences between them. The element of speed of reasoning was introduced by having maximum time limits for each item encountered. Fluid intelligence items written encompassed *odd man out* and *sequence* formats (see Table 5.2 examples).

Crystallised intelligence items were written to require verbal ability and sequential reasoning. The Victoria University of Wellington corpus of spoken New Zealand English guided the difficulty level of crystallised intelligence items and distractors. To design items varying in difficulty and discriminated across respondents, solutions and distractors spanned the frequency range of 2 to 98. Item distractors of similar frequency were written for each of these items to increase the likelihood of equal opportunity endorsement. Crystallised intelligence items written include the following formats: *synonyms* (e.g., the meaning of anticipated is closest to? 1) expected 2) normal 3) accepted 4) predictable 5) customary), *antonyms* (e.g., the meaning of gregarious is furthest from? 1) friendly 2) grumpy 3) unhappy 4) unsocial 5) unstable), *analogies* (e.g., queen is to prince as son is to? 1) cousin 2) sibling 3) father 4) mother 5) monarch), *classification* (e.g., which of the following is the odd one out? 1) leaf 2) branch 3) bark 4) shoot 5) tree), and *vocabulary* (e.g., an eddy is a? 1) small stream 2) cave dweller 3) whirlpool 4) small tree 5) small fossil).

Section 5.1.3: Procedures (Study 2)

All of those who responded to the invitation to participate encountered given information on its purpose and parameters. They were also required to click an "agree" button on the consent page before progressing to the assessments. All information provided to participants is reported in Appendix B.

If potential participants agreed to take part in the research they were directed to a "biographical information" page. This page asked for voluntary information on respondent demographics. The information sought and response options available are detailed in Table 5.1 in the "participants" section of this chapter. The demographic options provided were based on New Zealand Census categories. Participants were then able to move to a "general instructions" page. The information provided reiterated the structure of the research and provided more specific information on what would be encountered and how to complete the assessments.

More specific instructions and details were then provided to participants prior to each assessment section. These instructions made reference to both speed and accuracy, which is consistent with test administration that involves the collection of response time data (van Breukelen, 2005; Glickman et al., 2005). Each participant who completed the pilot were taken through the sections in the following order: crystallised intelligence (referred to as language ability), fluid intelligence (referred to as problem solving ability), and personality.

Although the sections were presented to participants in the same order, item presentation within each section was randomised. This randomisation was intended increase the likelihood that sufficient data would be collected for each item despite potential participant dropout rates. It was also done to reduce the likelihood that item order would artificially influence participant response choice (Ortner, 2008). Participants were presented with five response options for each item encountered. A box in the top right hand corner of the screen counted down from 50 seconds for each item. If the participant had not chosen a response within 40 seconds the countdown for the final ten seconds would flash in red. The use of such a response signal is consistent with common procedures within experimental psychology (Glickman et al., 2005). It is also consistent with recommendations by Kong, Wise, and Bhola (2007) to reduce rapid guessing behaviour through facilitating respondent time management. Participants were also instructed to guess the correct answer for items they could not solve. This is consistent with other participant instructions provided in RT related research (Schnipke & Scrams, 1997).

The response option selected and the number of seconds taken to make this choice was recorded. The choice to record RT in one second increments is consistent with other psychometric and educational test research (Schnipke & Scrams, 1997; Verbic, 2010; Wang & Hanson, 2005). Once participants had completed the assessments they were thanked for their participation and given the opportunity to provide anonymous qualitative feedback on their experience of the assessments. The analyses undertaken with collected data are detailed in the following section.

Section 5.1.4: Analysis (Study 2)

The data sets received included item by item responses, response latencies (in seconds), and demographic information per participant. The fluid and crystallised intelligence response options were then recoded as correct/incorrect choices. The personality response options were also recoded to score one through five depending on strength of endorsement (five equating with strong trait endorsement). This required the reverse scoring of items providing contrary indication of trait possession. An additional dichotomous recoding of personality responses was also undertaken to facilitate point-biserial correlations of corrected item total correlations (Henrysson, 1971). This was done by recoding one and two (strongly disagree and disagree) as zero/non-endorsement and four and five (agree and strongly agree) as one/endorse. Response option three (neutral/uncertain) was coded as a non-response.

Prior to conducting the core analysis detailed in the results section, distributions of the measures were examined. These revealed a small proportion of extremely fast responses. These were considered indicative of responses prior to processing item content. The values in this tail of the latency distribution are potentially problematic as they are likely to distort means and inflate variances (Greenwald, McGhee, & Schwartz, 1998; Schnipke & Scrams, 1997). On this basis responses of two seconds or less were excluded from analyses for ability items and participants with more than 80% of responses occurring in less than two seconds across any scale were removed from the data set. The use of this two second threshold is consistent with recommendations by Wise and DeMars (2006) and Kong, Wise, and Bhola (2007). It is also common to trim response times longer than 2.5 standard deviations about the mean (Glickman, Gray, & Morales, 2005), but the risk of obscuring true distributions within this investigation's data outweights any likely benefits of such trimming (Craigmile, Peruggia, & van Zandt, 2010). Cases were also excluded from subsequent analysis if they had failed to respond to ten percent or more of items on any of the assessments. These criteria for exclusion reduced the data set from 524 to 317 cases.

In line with recommendations, the psychometric properties of the assessments constructed were examined using a variety of analyses (Cohen & Swerdlik, 2002). This use of multiple analytic approaches is intended to reduce the likelihood that single estimates with inaccurately influence assessments of reliability and validity. The first item parameter detailed examines item discrimination power. Item discrimination power is assessed within both Classical Test Theory and Item Resposne Theory analyses. Both methods of analyses are intended to ensure that the items retained are the items that actually increase the consistency of what an assessment is measuring. For the CTT-based analyses point-biserial corrected item total correlations were calculated as a primary measure of item discrimination power (Kline, 2000). This item total correlation provides an indication of how strongly related a respondent's answer to an item is to their overall score on the relevant scale. A minimum point-biserial value of .15 is sometimes recommended, but the highest quality items are likely to have point-biserials above .25 (Varma, 2006). In line with this, most researchers set minimum correlation thresholds between .2 and .3 (Everitt, 2002; Field, 2005). To promote item quality items with correlations rounding to .25 or above in this investigation were retained for subsequent analysis. Item discrimination power was further tested by calculating each item's *discrimination index* (*d*). The item discrimination power index captures each item's ability to differentiate between respondents at the high and low sub-range of the construct measure (Choen & Swerdlik, 2010; Matlock-Hetzel, 1997). This investigation separated these top and bottom scorers via Nunnally's (1972) recommendation to use 25% for the upper and lower groups. According to Ebel and Frisbie (1986) item discrimination index values of .40 and greater are very good, .30 to .39 are reasonably good, .20 to .29 are marginal, and below .19 is poor. On this basis items with *d* values below .2 were excluded from subsequent analyses. The discrimination index was calculated using Salkind's (2006) formula:

$$d = \frac{Nu - Nl}{(0.5)\, Nt}$$

Where

d = discrimination value
Nu = number of correct responses in the upper quartile
Nl = number of correct responses in the lower quartile
Nt = total number of responses to the item

Ability item analysis also requires an estimation of item difficulty. This is to ensure that the items used will provide a spread of data and are more likely to result in a standard distribution. The most common index of item difficulty is the proportion of respondents that answered correctly (Kline, 2000). Varma (2006) suggests the reliability of assessments is likely to be highest when proportions correct are spread across the whole 0.0 to 1.0 range, with the greatest proportion around 0.50. Kline (2000) recommends that item difficulty be restricted from .2 to .8. This is increases the likelihood that items less than 20% of respondents will get right or wrong are excluded. A decision was made in this investigation to apply a proportions correct range of .1 to .9. This is in line with Leong and Austin's (1996) recommendations and makes it more likely that even those in the well below average range on the assessment will answer some items correctly.

The natural logarithm of the correct to incorrect ratio (logit *p*) serves as another index of item difficulty (Brown & White, 2009). According to Killen (2001) logit *p* is a closely related measure to the proportion of respondents that answer an item correctly. The rationale for its inclusion as an index of difficulty for crystallised and fluid intelligence items in this investigation is its ability to provide a statistic with satisfactory stability of variance across all intervals of item difficulty.

Distractor analysis is another fundamental component of ability test construction. Distractor analysis is intended to increase the likelihood that those answering items correctly are doing so because they have the required level of ability to do so, not because the alternative answers are not sufficiently distracting. A good distractor is a response option that is more likely to be chosen by low than high ability test takers (Haladyna, Downing, & Rodriguez, 2002). On this basis the functionality of distractors can be calculated by enumerating the percentile of upper and lower ability group members who choose each distractor for each item. According to Allen and Yen (1979) similar results will be found for upper and lower ability groups with boundary lines of any percentile between 25 and 33. On this basis the quartile-based upper and lower group separations used during item discrimination analysis were retained. Establishing the factors structure underlying items is a common step in test design once the discrimination, difficultly, and distractor success of items has been explored (Haladyna et al., 2002).

According to Kline (1994) factor analysis's ability to explore factor structure and ensure items selected only load on one factor makes factor analysis a crucial aspect of test construction. Kline (1994) suggests that exploratory factor analysis (EFA) is superior to confirmatory factor analysis (CFA) in test construction. He also suggests it is more consistent with the factor structure seeking rather than confirming examination of personality domains discussed in the literature review. Furthermore, it is an appropriate choice based on the relatively small participant group size employed (DeCoster, 1998). Research by Raubenheimer (2004) further supports the use of EFA to maximise scale reliability and validity. For these reasons the current investigation used EFA.

The first step EFA involved determining how many factors to extract. The most commonly recommended approach to this is to conduct a *Principal Components Analysis* (PCA) (e.g., Child, 2006; Kline, 1994, 2000). The criteria guiding the number of components a PCA has extracted are often unclear (Fabrigar, Wegener, MacCallum, & Strahan, 1999). For this reason the current investigation employed two of the most common methods for determining salient factors. The first of these was the Unity Rule (i.e., Kaiser's criterion), wherein only factors having eigenvalues greater than one are considered common factors. However, a criticism of the Unity Rule relevant to the current investigation is that too many factors meet this criterion when there are a large number of variables and low communalities (Cattell, 1978; Gorsuch, 1983). On this basis scree tests (Cattell 1966, 1978) were given the most weight in determining the number of salient factors for extraction. Interpretability served as a further guiding criterion for extraction. Interpretability refers to whether or not the factors are actually interpretable. In other words, do the items appear to be reasonably and clearly indicative of an underlying factor?

The final step in this initial stage of factor number identification required identifying significant factor loadings. This investigation employed the most prominent method in current use – selecting loadings of negative or positive .3 or greater. According to Child (2006) the minimum number of variables required to define a factor is three.

After determining the number of salient factors to extract via PCA a factor analysis aimed at detecting theoretically relevant underlying structures within the data is required. There is a variety of different methods of factor analysis. While the output of these methods is generally very similar, it is very important to compare results across multiple methods of extraction (Browne, 2001; Child, 2006). All factor analyses and factor rotations within this investigation used both *Principal Axis Factoring* (PAF) and *Maximum Likelihood Factor Extraction* (ML) for comparative purposes. The primary difference between these methods of extraction is ML's assumption that data are multivariate normal (Costello & Osborne, 2005). Although ML was undertaken for comparative purposes, a growing distrust of assessment score normality (Micceri, 1989) means PAF results are given most weight.

Many different types of rotation can be undertaken. These fall into the two basic categories of orthogonal and oblique rotation. Orthogonal rotation assumes no correlations between factors while oblique allows for such associations. Orthogonal rotation is considered the most interpretable category of rotation, but oblique is considered the most realistic regarding likely correlations between factors (Browne, 2001; Costello & Osborne, 2005). The results of rotations can vary considerably depending on which rotation method is employed (Sass & Schmitt, 2010). According to Kline (1994) both orthogonal and oblique rotations should be attempted. If they the results differ substantially then the oblique rotation becomes the focus. If the results are sufficiently similar, then the easier interpretability of the orthogonal rotation makes it the focus of most attention. Both oblique and orthogonal rotations were undertaken.

The most common and widely recommended method of orthogonal rotation is *varimax. Direct oblimin* is considered the best choice for obtaining a simple structure via oblique rotation (Costello & Osborne, 2005; Kline, 1994). The importance of a simple structure is outlined by Thurstone (1947). Kline (1994) defines a simple structure as a factor matrix wherein each factor has only a few high loadings. This investigation uses direct oblimin and varimax when undertaking oblique and orthogonal rotations respectively.

Oblique rotations provide information on pattern and structure matrices. The pattern matrix contains the relative weights for factor scores, while the structure matrix details the simple correlations of variables with factors. There is considerable disagreement as to which of these matrices should be the focus of interpretation during an oblique rotation. Kline (1994) and Pett, Lackey, and Sullivan

(2003) all argue for the structure matrix. By contrast, Ho (2006) and Tabachnick and Fidell (2001) all argue in favour of the pattern matrix. This investigation follows Child's (2006) advice to look at the pattern matrix first, then look at the structure matrix, then think about the implications/meaning of any differences observed. Another important consideration within test design is the precision of items that load upon the identified factors and related scales.

Measurement precision is not consistent across the items within a scale of measurement. According to Embretson and Reise (2000) CTT designed assessments are generally better at distinguishing between those with moderate ability levels than high- and low-scoring test-takers. The use of item response theory in test design is able to extend the CTT concept of reliability as a single index to a function called the *information function*. The information function within IRT is the inverse of the conditional observed score standard error at any given ability score. Item response theory also facilitates an increase in the construct validity of tests through the more accurate identification of items measuring (or not measuring) the intended to be measured skills (Ackerman, 1992). For these reasons, fluid and crystallised intelligence items were subjected to an item response theory evaluation and analysis. The decision to only employ IRT in the analysis of ability items is consistent with Kline's (1992) view that IRT is most suited to tests of attainment or ability (*cf.* Embretson & Reise, 2000; Harvey & Hammer, 1999). It is also consistent with the paucity of literature providing psychological meaning to the interpretation of IRT parameters for non-ability tests (Hubley, Wu, & Zumbo, 2009).

Investigations examining the fit between dichotomously scored unidimensional cognitive ability data and IRT models have suggested the three parameter (3PL) (Birnbaum, 1968) is an excellent selection (Chernyshenko, Stark, & Drasgow, 2008). Furthermore, respondents within the current investigation were unable to skip items and were instructed to guess if unsure of the correct response. This makes the inclusion of the *c*-parameter (likelihood of guessing the correct response option) an advantage over simpler one- and two-parameter logistic models. On this basis the *eirt - Item Response Theory Assistant for Excel* (Germain, Valois, & Belkacem, 2007) was used to make 3PL item parameter estimates for both fluid and crystallised intelligence scales. These calculations employed the Baye's modal estimator method (BME) with expected *a posteriori* (EAP) estimation (Bock & Mislevy, 1982).

Refined fluid and crystallised ability total scores were then weighted in line with Samejima's (1969) graded response (GR) model. The GR model suggests that the items better able to discriminate among examinees should be given higher weights as they will contribute more information to the final examinee score estimates. Lord (1980) also makes a strong case for using IRT to create weights for ability items. In line with Samejima's GR model, Lord demonstrated that the optimal weights for dichotomously scored items equal the slope (*a*) parameters in the IRT model. As is consistent with Comrey's (1988) recommendations, personality items were refined via EFA, but not weighted.

Statistical analyses were also undertaken to explore the degree of shared variance among variables. The strength of association measure used was Pearson-Product Moment correlation coefficients (i.e., Person's *r*). Pearson's *r* is the most appropriate index of shared variance with continuous and normally distributed data (Ferguson, 2009). As in the last chapter, Ferguson's (2009) recommendations for effect size magnitude were given most weight.

Forward stepwise multiple regressions were also calculated to provide additional information on the relationships among integrity, conscientiousness, and neuroticism. These regressions all utilised integrity as the criterion variable. Criterion choice was based upon previously hypothesised relationships and further explication of correlations observed in analyses of coefficients. Multiple regressions were performed because the information provided concerns the combined variance explained by a group predictors rather than the individual relationships previously detailed (Cohen et al., 2003).

Cohen's discrimination index (*d*) was also calculated to examine potential differences in mean scale scores between Māori and the broader participant group. The decision to use Cohen's *d* rather than IRT's potentially superior differential item functions calculations (Glickman, Seal, & Eisen, 2009; Hays, Morales, & Reise, 2000) was based on the fact the data fell far short of participant group size requirements for the latter (Harvey & Hammer, 1999; Stark & Chernyshenko, 2006). The significance of differences was calculated by an independent measures *t*-test.

Section 5.2: Results (Study 2)

Unless otherwise specified, this section details analyses utilising the statistical software package *PASW 18*. Statistical analyses were undertaken to explore item parameters, reliability, validity, and test hypotheses. Where relevant to analysis only results with *p*-levels .007 or below were deemed significant. This *p*-level was determined by making a Bonferroni correction (i.e., dividing the probability value [.05] by the number of tests to be conducted [7]). The minimal participant group size necessary to achieve a power-level of .8 was identified as N = 115. This participant group size was exceeded within the majority of calculations.

The first section of the results details item-level analysis. The second tentatively looks at reliability and validity. The third uses a three-parameter logistic (3PL) Item Response Theory (IRT) model to more clearly establish ability item parameters. Each of these sections and analyses serve as a hurdle for the inclusion of items in subsequent analyses and sections. The decision to exclude items not meeting selection criteria for subsequent analysis is based upon brevity, Raubenheimer's (2004) findings, and Nunnally's (1978) recommendations. The fourth section investigates Chapter 3's research objectives. The data in this section met the required statistical assumptions to perform multiple regressions and correlations (Cohen et al., 2003). Most relevantly, the data were normally distributed; displayed linearity and homoscedasticity; and measurement appeared reliable (Osborne & Waters, 2002).

A number of items met the criteria for discrimination and difficulty. The Cronbach's Alpha (α) scores for these items suggested that aside from the crystallised ability scale, all scales exceeded the recommended standard of .7 (Eatwell, 1997). Scale reliability was also supported across gender and ethnic groups. Specific Alpha information is contained in Tables 5.17 through 5.19. Exploratory factor analysis also established and supported the underlying factor structure for both ability and personality items. Table 5.15 and Table 5.16 detail these findings. Inter-scale correlation coefficients reported in Table 5.25 tentatively support the discriminant validity of assessments.

Item response theory clarified item parameters for all of the ability items that had progressed through the CTT retention criteria (see Table 5.20 and Table 5.21). The scope and precision of ability scales is illustrated via item information function curves in Figure 5.7 and Figure 5.8.

Once scale and item parameters were established it was possible to investigate the research objectives of the dissertation. Results from Study 1 were generally replicated for the first and second research objectives. Results relating to the third research objective found the length of time taken to answer ability assessments was related to respondent ability. It was also discovered that differences in conscientiousness meant respondents take longer to complete fluid ability assessments. A similar result was observed for one neuroticism facet and the length of time taken to complete fluid ability questions. Section 4.2.1 reports results relating to the item-level analysis undertaken as an antecedent to investigating the research objective.

Section 5.2.1: Item-level Analysis

This section contains information concerning the basic descriptive statistics of item responses and accompanying response latencies. It also details results pertaining to item discrimination, item difficulty, and item distractor analysis.

Section 5.2.1.1: Descriptive Statistics

Tables 5.3 through to 5.8 detail item descriptive statistics. These include scale information on means and standard deviations for both scores and response time latencies. The items in Table 5.3 were dichotomously coded. Table 5.3 shows that the average response latencies for the 30 crystallised intelligence items ranged from 6.54 to 17 seconds.

Table 5.3

Descriptive Statistics for Crystallised Intelligence Items

		Score				Response Latency (Sec.)			
Item	N	Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	315	0	1	.48	.50	3	39	10.41	5.84
2	314	0	1	.92	.27	3	28	8.36	4.62
3	314	0	1	.29	.45	3	49	13.79	7.20
4	315	0	1	.52	.50	3	45	11.12	6.61
5	316	0	1	.95	.22	3	44	7.50	4.46
6	315	0	1	.85	.35	3	35	10.03	5.18
7	314	0	1	.68	.46	3	50	17.20	9.97
8	310	0	1	.54	.49	3	48	16.55	8.85
9	317	0	1	.88	.32	3	45	10.48	6.13
10	313	0	1	.54	.49	3	45	15.05	7.52
11	314	0	1	.97	.15	3	47	9.67	5.02
12	312	0	1	.98	.13	3	27	7.80	3.65
13	313	0	1	.97	.17	3	47	6.97	3.95
14	316	0	1	.73	.44	3	33	9.14	4.23
15	314	0	1	.84	.36	3	34	11.31	5.61
16	315	0	1	.23	.41	3	41	12.95	6.48
17	314	0	1	.93	.25	3	45	12.58	6.24
18	313	0	1	.39	.48	3	48	15.87	9.87
19	317	0	1	.43	.49	3	38	12.28	6.17
20	313	0	1	.51	.50	3	48	12.90	8.16
21	313	0	1	.88	.32	3	49	10.01	6.25
22	316	0	1	.63	.48	3	45	11.82	6.14
23	314	0	1	.50	.50	3	48	10.25	6.09
24	315	0	1	.79	.40	3	50	12.14	7.29
25	316	0	1	.66	.47	3	40	12.24	6.53
26	316	0	1	.45	.49	3	41	9.04	5.99
27	312	0	1	.43	.49	3	50	12.51	8.95
28	317	0	1	.98	.13	3	36	6.65	3.54
29	314	0	1	.81	.39	3	28	7.66	4.16
30	317	0	1	.70	.46	3	30	8.19	4.51

Note. Listwise deletion of missing data.

Table 5.4 shows that fluid intelligence items were also dichotomously coded. The average time taken for the 30 items detailed below ranges from 14.02 to 32.77 seconds.

Table 5.4

Descriptive Statistics for Fluid Intelligence Items

Item	N	Score				Response Latency (Sec.)			
		Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	315	0	1	.85	.35	3	50	14.18	7.80
2	308	0	1	.19	.39	3	50	27.90	10.96
3	304	0	1	.33	.47	3	50	28.11	12.77
4	314	0	1	.65	.47	3	50	21.81	10.80
5	315	0	1	.64	.48	3	50	22.35	12.28
6	314	0	1	.08	.26	3	50	22.93	11.87
7	313	0	1	.75	.43	3	49	14.49	8.58
8	310	0	1	.53	.50	3	50	23.27	11.31
9	287	0	1	.20	.39	3	50	32.87	11.98
10	314	0	1	.57	.49	3	50	22.52	10.80
11	299	0	1	.57	.49	3	50	25.68	11.39
12	316	0	1	.72	.45	3	46	14.45	8.00
13	315	0	1	.47	.50	3	50	15.85	8.87
14	312	0	1	.15	.35	3	43	13.47	7.57
15	310	0	1	.07	.26	3	50	22.07	11.88
16	311	0	1	.59	.49	3	50	22.73	11.91
17	316	0	1	.91	.28	3	50	12.99	7.58
18	316	0	1	.69	.46	3	48	14.45	8.06
19	312	0	1	.64	.48	3	50	20.00	9.85
20	314	0	1	.72	.44	3	50	23.63	10.55
21	310	0	1	.65	.47	3	49	22.90	10.67
22	314	0	1	.82	.38	3	49	14.50	7.69
23	314	0	1	.71	.45	3	50	15.19	7.51
24	303	0	1	.51	.50	3	50	28.12	12.61
25	311	0	1	.66	.47	3	50	19.83	9.79
26	309	0	1	.74	.44	3	50	22.00	10.83
27	315	0	1	.09	.28	3	50	15.77	9.00
28	309	0	1	.12	.32	3	49	19.02	9.89
29	289	0	1	.28	.45	3	50	32.35	12.44
30	305	0	1	.52	.50	3	50	26.05	12.40

Note. Listwise deletion of missing data.

Table 5.5 shows that integrity item answers were coded on a five-point basis. The average time taken for the 40 items detailed below ranges from 3.80 to 10.95 seconds.

Table 5.5

Descriptive Statistics for Integrity Items

		Score				Response Latency (Sec.)			
Item	N	Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	317	1	5	2.94	1.06	1	47	10.08	6.20
2	314	1	5	2.12	1.04	1	49	9.05	5.44
3	317	1	5	2.66	1.06	1	38	5.46	4.23
4	316	1	5	2.05	0.94	1	24	6.84	3.05
5	312	1	5	3.87	0.85	2	29	6.77	3.63
6	316	1	4	1.92	0.69	1	36	6.45	4.15
7	316	1	5	1.98	1.01	1	21	6.25	3.11
8	315	1	5	3.18	1.02	2	40	5.92	4.03
9	314	1	5	2.64	1.17	1	24	6.74	3.51
10	316	1	5	2.48	1.12	1	38	6.31	4.00
11	317	1	5	3.12	1.12	1	50	5.46	3.86
12	317	1	5	3.43	1.04	1	46	5.13	3.61
13	315	1	5	3.19	1.33	1	38	5.93	3.72
14	317	1	4	1.47	0.58	2	48	10.95	5.85
15	317	1	5	2.42	0.87	1	15	3.80	2.30
16	316	1	5	2.05	0.97	1	31	5.59	3.60
17	317	1	5	2.11	0.91	1	42	7.81	5.18
18	314	1	5	2.65	1.08	1	40	6.36	4.31
19	315	1	5	2.37	1.11	1	48	7.39	4.82
20	315	1	5	1.82	0.96	1	50	8.43	4.95
21	315	1	5	2.65	0.98	1	50	8.23	5.25
22	317	1	4	1.58	0.68	1	50	5.32	3.78
23	314	1	5	1.95	0.89	1	37	6.60	4.32
24	315	1	5	2.10	0.97	1	20	5.84	2.72
25	315	1	5	1.85	0.81	2	24	5.22	2.81
26	315	1	5	3.23	0.98	1	46	6.78	4.84
27	312	1	5	1.62	0.75	2	29	5.85	3.64
28	315	1	5	2.71	1.14	2	40	7.10	4.61
29	316	1	4	1.73	0.81	2	43	7.22	4.23
30	315	1	5	2.93	1.07	2	45	7.39	4.46
31	312	1	5	3.24	1.05	1	44	6.63	4.43
32	317	1	5	2.35	0.98	1	41	6.48	3.95
33	317	1	5	2.10	0.90	1	36	7.32	4.45
34	313	1	5	1.93	0.82	1	46	8.15	4.81
35	315	1	5	2.41	0.90	1	33	6.33	3.82
36	314	1	4	1.84	0.79	1	40	6.11	3.98
37	317	1	5	2.49	1.10	2	28	6.17	3.66
38	314	1	5	2.89	1.21	1	36	6.38	3.94
39	315	1	5	2.59	1.19	1	30	5.31	3.26
40	314	1	5	3.53	1.07	2	37	8.52	4.62

Note. Listwise deletion of missing data.

Table 5.6 and Table 5.7 show that conscientiousness and neuroticism item answers were coded on a five-point basis. The average time taken for conscientiousness items ranged from 3.22 to 8.33 seconds. The average time taken for neuroticism items ranged from 3.19 to 7.35 seconds.

Table 5.6

Descriptive Statistics for Conscientiousness Items

		Score				Response Latency (Sec.)			
Item	N	Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	315	1	5	2.71	0.98	1	46	7.24	4.77
2	317	1	5	2.37	0.91	1	42	3.87	3.46
3	317	1	5	3.80	0.93	1	27	3.22	2.22
4	316	1	5	3.60	1.09	1	27	5.60	3.27
5	315	1	5	3.86	0.88	2	41	5.96	3.82
6	315	1	5	3.76	1.06	1	16	3.69	2.00
7	317	1	5	3.91	0.92	1	24	6.54	2.93
8	316	1	5	3.07	1.18	1	27	6.90	3.40
9	315	1	5	2.88	1.11	2	35	6.23	3.45
10	315	1	5	3.10	1.13	1	42	5.56	3.46
11	314	1	5	3.89	0.78	1	44	4.24	3.16
12	315	1	5	3.26	1.08	1	47	7.12	4.77
13	315	1	5	2.94	1.03	2	31	5.05	3.01
14	316	1	5	3.02	1.14	1	34	5.50	3.42
15	315	1	5	3.50	1.07	1	26	5.12	2.54
16	317	1	5	3.61	1.02	1	22	5.45	3.23
17	314	1	5	3.80	0.95	1	48	5.55	4.17
18	315	1	5	2.76	1.02	1	23	6.59	3.45
19	317	1	5	2.97	1.08	1	40	6.34	3.60
20	316	1	5	2.34	0.84	1	48	4.03	3.60
21	316	1	5	3.41	0.98	1	35	4.91	3.70
22	317	1	5	4.25	0.84	1	19	4.28	2.07
23	317	1	5	3.47	1.04	1	16	4.26	2.30
24	316	1	5	3.37	0.97	1	38	7.66	4.47
25	316	1	5	4.04	0.77	1	20	4.45	2.44
26	316	1	5	2.80	1.00	1	50	5.86	4.25
27	317	1	5	3.61	1.16	1	27	5.08	2.72
28	316	1	5	3.94	0.86	1	48	5.11	3.47
29	315	1	5	3.56	0.96	1	18	4.98	2.61
30	315	1	5	3.45	1.09	1	50	4.49	3.52
31	316	1	5	3.64	0.99	1	25	4.49	2.94
32	317	1	5	3.23	1.04	1	48	4.97	4.33
33	316	1	5	3.96	0.68	1	49	4.70	3.93
34	313	1	5	2.92	1.09	1	49	7.29	5.46
35	313	1	5	3.85	0.79	1	24	4.34	2.49
36	317	1	5	3.70	1.09	1	24	4.10	2.61
37	317	1	5	4.03	0.84	1	38	5.09	3.52
38	315	1	5	2.79	1.01	1	50	8.31	6.10
39	316	1	5	3.08	1.06	1	47	8.33	5.49
40	316	1	5	3.34	0.98	1	17	5.40	2.75
41	316	1	5	3.55	0.96	1	49	8.18	5.86
42	314	1	5	3.87	0.98	1	50	6.77	4.15
43	316	1	5	3.42	0.89	2	29	5.40	3.25
44	317	1	5	3.63	0.97	1	46	5.88	4.50
45	316	1	5	2.91	1.02	2	36	7.01	4.69
46	316	1	5	2.84	1.11	1	28	6.30	3.32
47	317	1	5	2.99	1.04	2	50	7.66	5.27
48	316	1	5	3.57	1.09	1	13	3.49	2.03
49	315	1	5	4.17	0.72	2	31	5.16	2.99
50	317	1	5	4.10	0.83	1	15	4.16	2.11

Note. Listwise deletion of missing data.

Table 5.7

Descriptive Statistics for Neuroticism Items

Item	N	Score				Response Latency (Sec.)			
		Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	317	1	5	1.94	0.78	1	29	3.19	2.68
2	316	1	5	3.31	1.09	2	29	5.34	3.68
3	317	1	5	3.50	0.98	1	49	5.43	4.22
4	317	1	5	1.97	0.76	1	22	3.47	2.45
5	317	1	5	3.12	1.10	1	20	4.00	2.50
6	316	1	5	2.33	1.04	1	20	4.75	2.88
7	316	1	5	2.92	1.03	1	36	5.57	3.91
8	316	1	5	3.73	0.86	1	48	4.10	3.47
9	317	1	5	2.14	0.90	1	34	4.99	3.01
10	311	1	5	2.68	0.93	2	45	5.58	4.22
11	316	1	5	2.50	1.04	1	21	3.79	2.50
12	316	1	5	2.59	1.09	1	47	5.14	4.09
13	315	1	5	2.83	1.07	2	49	6.04	3.98
14	315	1	5	2.57	1.14	1	22	4.88	3.13
15	316	1	5	2.60	1.01	1	21	3.91	2.52
16	316	1	5	2.09	0.89	1	46	5.57	3.90
17	317	1	5	3.54	0.91	2	41	5.48	4.25
18	314	1	5	3.21	1.06	1	23	3.99	2.70
19	315	1	5	2.76	1.14	1	24	5.02	3.13
20	316	1	5	2.33	1.00	1	30	5.64	3.40
21	316	1	5	2.52	0.98	2	40	7.35	4.62
22	314	1	5	2.18	0.88	1	14	3.85	1.93
23	316	1	5	2.06	0.89	1	33	3.88	2.84
24	316	1	5	3.02	1.08	1	33	5.22	3.85
25	317	1	5	2.99	1.14	1	46	6.15	3.79
26	315	1	5	3.31	1.12	1	28	4.73	2.91
27	315	1	5	3.56	0.99	1	25	5.12	2.86
28	317	1	5	3.39	1.00	1	30	4.59	3.48
29	317	1	5	3.36	1.05	1	32	4.97	3.18
30	317	1	5	3.31	1.07	1	32	4.51	3.14
31	316	1	5	2.69	1.01	1	21	5.23	2.90
32	317	1	5	3.60	0.97	1	43	5.22	4.36
33	317	1	5	2.74	0.98	1	26	5.59	3.34
34	315	1	5	3.07	1.08	1	26	6.08	3.65
35	317	1	5	2.21	1.12	1	27	4.57	3.07
36	315	1	5	2.94	1.18	1	27	3.85	2.75
37	316	1	5	2.21	0.78	1	50	4.37	3.51
38	315	1	5	2.81	1.34	1	21	4.91	2.57
39	314	1	5	3.33	1.07	1	33	6.37	3.81
40	314	1	5	2.40	1.15	1	30	6.99	4.00
41	316	1	5	2.20	1.01	1	46	4.59	3.53
42	316	1	5	3.13	1.19	1	43	6.02	4.30
43	317	1	5	3.47	0.97	1	41	4.04	3.36
44	316	1	5	3.11	1.07	1	23	5.35	3.09
45	316	1	5	3.39	1.07	1	26	4.34	2.95
46	316	1	5	2.71	1.17	1	42	5.26	4.32
47	317	1	5	3.19	1.09	2	30	5.56	3.29
48	317	1	5	2.74	1.09	2	34	6.80	4.16
49	314	1	5	2.73	1.16	1	33	5.97	3.76
50	316	1	5	2.29	0.89	1	15	3.59	2.23

Note. Listwise deletion of missing data.

Dissimulation item answers reported in Table 5.8 were coded on a five-point basis. The average time taken for the 20 items ranged from 3.86 to 6.30 seconds.

Table 5.8

Descriptive Statistics for Dissimulation Items

		Score				Response Latency (Sec.)			
Item	N	Min.	Max.	Mean	Std. D.	Min.	Max.	Mean	Std. D.
1	317	1	5	2.34	0.94	1	31	4.59	3.35
2	317	1	5	1.86	0.62	1	29	4.39	2.73
3	316	1	5	2.52	0.95	1	43	5.00	3.83
4	316	1	5	2.12	0.93	1	32	4.05	2.86
5	317	1	5	2.43	1.06	1	23	4.35	2.97
6	317	1	5	2.58	1.07	1	20	4.44	2.54
7	317	1	5	3.03	1.17	1	23	5.00	2.89
8	316	1	5	3.05	1.00	1	20	4.13	2.61
9	317	1	5	2.69	1.02	1	33	4.58	3.21
10	314	1	5	2.97	1.03	1	22	5.04	2.61
11	316	1	5	2.64	1.06	2	36	6.30	4.02
12	317	1	5	2.81	1.00	1	23	4.32	2.75
13	317	1	5	3.38	1.01	1	27	3.89	2.68
14	316	1	5	2.24	0.75	1	43	5.36	4.22
15	315	1	5	2.81	0.99	2	30	5.30	3.30
16	315	1	5	3.41	1.07	1	21	4.32	2.65
17	315	1	5	2.39	0.91	1	43	4.60	4.24
18	316	1	5	2.39	0.97	1	22	5.20	2.90
19	317	1	5	1.79	0.77	1	25	4.42	2.52
20	316	2	5	4.22	0.75	1	20	3.86	2.32

Note. Listwise deletion of missing data.

Section 5.2.1.2: Item Discrimination Power

As mentioned in Section 5.1.4, item discrimination power concerns how well respondent answers differentiate the relative strength of their affiliation with a construct measured. The corrected item-total correlations (CITC) detailed in Tables 5.9 through 5.11 are point-biserial correlations between right/wrong item scores and total scale scores. These CITCs provide an indication of how well responses to an item discriminate between overall scale scores. Items with point-biserials rounding to $r \geq .25$ met the threshold for retention (Varma, 2006). Tables 5.9 through 5.11 also detail how well items discriminated between respondents who score/endorse scales in the top versus bottom one-fourth (or one-quarter) of respondents according to construct measures. The *observed frequencies* in Table 5.9 indicate what percentage of the top one-quarter of respondents versus the bottom one-quarter answered the fluid and crystallised intelligence items correctly. The observed frequencies columns in

Table 5.10 and Table 5.11 instead detail what percentage of the top one-quarter and bottom one-quarter endorsed personality items. The discrimination index (*d* value) details exactly how well each item discriminated between those who scored high and low on the total test. Only items with *d* values of .2 or above are retained for subsequent analysis (Ebel & Frisbie, 1986).

Fifteen of the 30 crystallised intelligence items in Table 5.9 met the retention criteria of a CITC ≥ .25 and *d* value ≥ .2 (1, 3, 6, 14-16, 18-20, 22, 24-27, 29). These items ranged in CITC from *r* = .23 to *r* = .33 and had discrimination values between *d* = .22 to *d* = .58. Seventeen of the 30 fluid intelligence items in Table 5.9 also met the required criteria (1, 2, 4, 8-12, 16, 19-26). These items ranged in CITC from *r* = .24 to *r* = .55 and had discrimination values between *d* = .31 to *d* = .73.

Table 5.9

Crystallised and Fluid Intelligence Corrected Item-Total Correlations (CITC) and Discrimination Indices

	Crystallised Items					Fluid Items				
Items	CITC	N	*d*	L Quartile	U Quartile	CITC	N	*d*	L Quartile	U Quartile
1	**.33**	171	**0.44**	22 (24.7%)	60 (73.2%)	**.27**	186	**0.38**	60 (68.2%)	96 (98.0%)
2	.18	170	0.12	70 (79.5%)	81 (98.8%)	**.29**	181	**0.43**	03 (03.5%)	42 (44.2%)
3	**.26**	170	**0.35**	15 (17.0%)	45 (54.9%)	.11	176	0.27	21 (25.0%)	45 (48.9%)
4	-.03	171	0.14	39 (43.3%)	51 (63.0%)	**.34**	184	**0.59**	32 (36.8%)	87 (89.7%)
5	.17	172	0.03	77 (85.6%)	80 (97.6%)	.17	185	0.42	44 (50.0%)	83 (85.6%)
6	**.31**	170	**0.25**	57 (64.8%)	79 (96.3%)	.11	184	0.05	04 (04.5%)	09 (09.4%)
7	.08	170	0.09	53 (59.6%)	61 (75.3%)	-.16	184	0.04	67 (76.1%)	71 (74.0%)
8	.12	166	0.22	36 (42.4%)	55 (67.9%)	**.24**	181	**0.57**	22 (25.9%)	74 (77.1%)
9	.11	172	0.07	72 (80.0%)	78 (95.1%)	**.33**	166	**0.31**	04 (05.1%)	30 (34.5%)
10	.03	169	0.14	39 (44.8%)	51 (62.2%)	**.42**	184	**0.68**	19 (21.8%)	82 (84.5%)
11	-.02	169	-0.07	85 (96.6%)	79 (97.5%)	**.34**	179	**0.51**	28 (33.3%)	74 (77.9%)
12	.08	170	-0.02	84 (95.5%)	82 (100%)	**.36**	185	**0.59**	37 (42.5%)	92 (93.9%)
13	.12	172	0.00	82 (91.1%)	82 (100%)	.12	184	0.28	33 (37.5%)	59 (61.5%)
14	**.25**	171	**0.36**	43 (48.3%)	74 (90.2%)	.18	185	0.17	07 (08.0%)	23 (23.7%)
15	**.25**	171	**0.28**	54 (60.7%)	78 (95.1%)	.10	182	0.12	02 (02.4%)	13 (13.4%)
16	**.24**	170	**0.35**	08 (09.0%)	38 (46.9%)	**.31**	183	**0.50**	31 (35.6%)	77(80.2%)
17	.01	171	-0.01	80 (89.9%)	79 (96.3%)	.20	186	0.24	73 (83.0%)	96 (98.0%)
18	**.33**	171	**0.52**	10 (11.1%)	55 (67.9%)	.17	186	0.28	53 (60.2%)	79 (80.6%)
19	**.31**	172	**0.58**	13 (14.4%)	63 (76.8%)	**.55**	185	**0.73**	21 (23.9%)	89 (91.8%)
20	**.29**	169	**0.39**	32 (36.4%)	65 (80.2%)	**.31**	184	**0.52**	42 (48.3%)	90 (92.8%)
21	.20	170	0.11	68 (77.3%)	78 (95.1%)	**.37**	183	**0.53**	35 (40.7%)	84 (86.6%)
22	**.32**	171	**0.49**	30 (33.7%)	72 (87.8%)	**.51**	185	**0.53**	47 (53.4%)	96 (99.0%)
23	.21	170	0.47	23 (26.1%)	63 (76.8%)	**.43**	185	**0.59**	36 (41.4%)	91 (92.9%)
24	**.29**	171	**0.29**	54 (60.7%)	79 (96.3%)	**.36**	179	**0.60**	18 (21.7%)	72 (75.0%)
25	**.23**	172	**0.30**	44 (48.9%)	70 (85.4%)	**.27**	180	**0.51**	34 (39.5%)	80 (85.1%)
26	**.23**	172	**0.54**	16 (17.8%)	63 (76.8%)	**.43**	184	**0.59**	37 (43.0%)	92 (93.9%)
27	**.28**	169	**0.47**	19 (21.8%)	59 (72.0%)	.12	185	0.10	04 (04.5%)	14 (14.4%)
28	.17	172	-0.02	84 (93.3%)	82 (100%)	.13	183	0.18	01 (01.1%)	18 (18.8%)
29	**.23**	171	**0.22**	57 (64.0%)	76 (92.7%)	.20	174	0.39	07 (08.2%)	41 (46.1%)
30	.08	172	0.16	54 (60.0%)	68 (82.9%)	.19	180	0.42	28 (32.2%)	66 (71.0%)

Note. Listwise deletion of missing data. Total N for Crystallised CITC = 264. Total N for Fluid CITC = 169. Quartile groupings and *d* are discrimination indices.

Thirty-one of the 50 conscientiousness items in Table 5.10 met the dual item retention criteria (1, 3-5, 8, 10-12, 15-17, 19, 22-24, 26-32, 34-36, 41, 42, 45, 46, 48, 50). These items ranged in CITC from *r* = .22 to *r* = .60 and had discrimination values between *d* = .30 to *d* = .86. Forty-two of the 50 neuroticism items in Table 5.10 also met the required criteria (2, 3, 5-20, 22-26, 28, 30-36, 38-44, 46, 47, 49, 50). These items ranged in CITC from *r* = .27 to *r* = .56 and had discrimination values between *d* = .28 to *d* = .88.

Table 5.10

Item Discrimination Power for Conscientiousness and Neuroticism

	Conscientiousness					Neuroticism				
Item	CITC	N	*d*	L Quartile	U Quartile	CITC	N	*d*	L Quartile	U Quartile
1	**.40**	145	**0.64**	03 (04.3%)	50 (65.8%)	.17	149	0.13	03 (03.8%)	13 (18.6%)
2	.03	150	0.06	17 (25.0%)	22 (26.8%)	**.53**	158	**0.87**	11 (14.9%)	80 (95.2%)
3	**.60**	153	**0.79**	28 (43.8%)	89 (100%)	**.38**	151	**0.62**	28 (39.4%)	75 (93.8%)
4	**.51**	158	**0.72**	27 (39.1%)	84 (94.4%)	.29	147	0.19	01 (01.3%)	15 (22.4%)
5	**.32**	162	**0.43**	51 (69.9%)	86 (96.6%)	**.52**	160	**0.80**	13 (17.1%)	77 (91.7%)
6	.13	168	0.31	53 (67.9%)	79 (87.8%)	**.36**	155	**0.46**	05 (06.7%)	41 (51.2%)
7	.18	157	0.39	51 (72.9%)	82 (94.3%)	**.47**	145	**0.75**	08 (12.1%)	63 (79.7%)
8	**.35**	166	**0.59**	19 (25.3%)	68 (74.7%)	**.27**	151	**0.49**	41 (61.2%)	78 (92.9%)
9	.20	159	0.47	20 (28.6%)	58 (65.2%)	**.47**	155	**0.38**	02 (02.5%)	32 (43.2%)
10	**.38**	158	**0.62**	16 (22.2%)	65 (75.6%)	**.29**	141	**0.39**	09 (13.0%)	37 (51.4%)
11	**.37**	156	**0.55**	48 (73.8%)	91 (100%)	**.53**	159	**0.65**	03 (03.8%)	55 (69.6%)
12	**.38**	155	**0.65**	22 (31.4%)	73 (85.9%)	**.50**	159	**0.64**	06 (07.5%)	57 (72.2%)
13	.17	148	0.39	20 (29.4%)	49 (61.3%)	**.40**	153	**0.64**	08 (10.7%)	57 (73.1%)
14	.19	159	0.40	25 (34.2%)	57 (66.3%)	**.48**	156	**0.65**	03 (03.9%)	54 (68.4%)
15	**.37**	154	**0.66**	27 (40.9%)	78 (88.6%)	**.42**	151	**0.59**	03 (03.8%)	48 (65.8%)
16	**.59**	150	**0.86**	22 (36.1%)	87 (97.8%)	**.32**	154	**0.28**	02 (02.5%)	24 (32.0%)
17	**.46**	159	**0.56**	41 (60.3%)	86 (94.5%)	**.36**	158	**0.53**	35 (46.1%)	77 (93.9%)
18	.15	144	0.40	16 (26.7%)	45 (53.6%)	**.47**	147	**0.81**	12 (18.2%)	72 (88.9%)
19	**.45**	155	**0.77**	09 (12.5%)	69 (83.1%)	**.47**	160	**0.66**	11 (14.1%)	64 (78.0%)
20	.21	148	0.32	02 (02.9%)	26 (32.5%)	**.49**	148	**0.52**	00 (00.0%)	39 (54.9%)
21	-.10	155	-0.05	60 (83.3%)	56 (67.5%)	.15	146	0.28	11 (16.7%)	32 (40.0%)
22	**.24**	164	**0.30**	63 (85.1%)	88 (97.8%)	**.29**	142	**0.28**	03 (03.8%)	23 (35.9%)
23	**.47**	150	**0.80**	20 (32.8%)	80 (89.9%)	**.38**	155	**0.34**	00 (00.0%)	27 (36.0%)
24	**.24**	144	**0.47**	31 (49.2%)	65 (80.2%)	**.54**	150	**0.84**	05 (07.0%)	68 (86.1%)
25	.21	163	0.28	62 (82.7%)	85 (96.6%)	**.38**	156	**0.62**	13 (17.8%)	62 (74.7%)
26	**.29**	158	**0.54**	10 (13.9%)	53 (61.6%)	**.37**	149	**0.65**	24 (34.3%)	73 (92.4%)
27	**.51**	157	**0.75**	22 (32.4%)	81 (91.0%)	.05	150	0.14	52 (69.3%)	63 (84.0%)
28	**.22**	161	**0.37**	55(76.4%)	85 (95.5%)	**.31**	148	**0.55**	28 (42.4%)	69 (84.1%)
29	**.38**	145	**0.69**	28 (44.4%)	78 (95.1%)	-.15	142	-0.09	50 (69.4%)	43 (61.4%)
30	**.43**	159	**0.70**	25 (35.2%)	81 (92.0%)	**.45**	160	**0.65**	25 (32.9%)	77 (91.7%)
31	**.50**	152	**0.76**	29 (46.0%)	87 (97.8%)	**.36**	153	**0.51**	10 (13.9%)	49 (60.5%)
32	**.23**	153	**0.45**	30 (44.1%)	65 (76.5%)	**.42**	158	**0.63**	30 (41.1%)	80 (94.1%)
33	.01	156	0.16	67 (91.8%)	80 (96.4%)	**.56**	156	**0.73**	03 (04.0%)	60 (74.1%)
34	**.31**	142	**0.63**	11 (17.5%)	56 (70.9%)	**.49**	152	**0.77**	10 (13.9%)	69 (86.3%)
35	**.40**	159	**0.44**	51 (70.8%)	86 (98.9%)	**.52**	161	**0.57**	02 (02.6%)	48 (57.8%)
36	**.38**	164	**0.61**	35 (47.9%)	85 (93.4%)	**.42**	153	**0.64**	16 (21.3%)	65 (83.3%)
37	.11	164	0.14	67 (87.0%)	79 (90.8%)	.16	144	0.20	02 (02.9%)	17 (23.0%)
38	.06	146	0.16	23 (37.7%)	35 (41.2%)	**.41**	161	**0.58**	11 (13.9%)	58 (70.7%)
39	.14	146	0.45	24 (38.1%)	57 (68.7%)	**.53**	157	**0.82**	11 (15.1%)	76 (90.5%)
40	-.03	160	0.05	53 (73.6%)	57 (64.8%)	**.41**	158	**0.50**	05 (06.5%)	45 (55.6%)
41	**.34**	152	**0.64**	32 (51.6%)	81 (90.0%)	**.45**	156	**0.47**	03 (03.8%)	40 (51.9%)
42	**.45**	163	**0.58**	41 (56.2%)	89 (98.9%)	**.38**	158	**0.59**	20 (26.7%)	67 (80.7%)
43	.09	136	0.38	37 (64.9%)	63 (79.7%)	**.52**	156	**0.78**	22 (31.4%)	83 (96.5%)
44	.13	158	0.32	51 (70.8%)	77 (89.5%)	**.55**	152	**0.88**	09 (12.7%)	76 (93.8%)
45	**.47**	154	**0.74**	07 (09.7%)	64 (78.0%)	.18	157	0.40	39 (53.4%)	71 (84.5%)
46	**.26**	156	**0.47**	16 (21.6%)	53 (64.6%)	**.52**	155	**0.69**	07 (09.3%)	61 (76.3%)
47	.17	159	0.40	28 (40.6%)	60 (66.7%)	**.47**	151	**0.74**	14 (19.7%)	70 (87.5%)
48	**.29**	150	**0.52**	38 (61.3%)	77 (87.5%)	.12	143	0.19	19 (27.1%)	33 (45.2%)
49	.22	171	0.23	69 (86.3%)	89 (97.8%)	**.35**	140	**0.50**	11 (16.2%)	46 (63.9%)
50	**.47**	159	**0.46**	52 (75.4%)	89 (98.9%)	**.33**	144	**0.30**	03 (03.9%)	25 (37.3%)

Note. Listwise deletion of missing data. Totatl N for crystallised intelligence = 264. Total N for fluid CITC = 169. Quartile groupings and *d* are discrimination indices.

Table 5.11 shows 26 of the 40 integrity items met the required criteria (1-4, 7-12, 16-19, 21, 23, 24, 26, 30-32, 35, 37-40). These items ranged in CITC from *r* = .26 to *r* = .51 and had discrimination values between *d* = .25 to *d* = .85. Sixteen of the 20 dissimulation items also met the retention criteria (1, 3, 5-13, 15-18, 20). These items ranged in CITC from *r* = .28 to *r* = .47 and had discrimination values between *d* = .30 to *d* = .82.

Table 5.11

Item Discrimination Power for Integrity and Dissimulation

	Integrity					Dissimulation				
Item	CITC	N	*d*	L Quartile	U Quartile	CITC	N	*d*	L Quartile	U Quartile
1	**.51**	162	**0.85**	05 (07.0%)	74 (81.3%)	**.40**	151	**0.43**	01 (01.3%)	34 (47.2%)
2	**.26**	158	**0.30**	01 (01.2%)	25 (32.5%)	.20	155	0.10	00 (00.0%)	08 (10.8%)
3	**.49**	168	**0.70**	05 (06.3%)	64 (71.9%)	**.34**	151	**0.47**	04 (05.1%)	40 (54.8%)
4	**.30**	169	**0.27**	02 (02.3%)	25 (30.5%)	.18	163	0.19	03 (03.6%)	19 (23.8%)
5	.15	166	0.37	55 (71.4%)	86 (96.6%)	**.40**	158	**0.48**	02 (02.5%)	40 (51.3%)
6	.16	158	0.07	00 (00.0%)	06 (07.3%)	**.28**	161	**0.48**	05 (06.3%)	44 (54.3%)
7	**.47**	177	**0.41**	00 (00.0%)	37 (41.1%)	**.42**	153	**0.69**	10 (13.5%)	63 (79.7%)
8	**.34**	162	**0.65**	21 (29.2%)	74 (82.2%)	**.45**	143	**0.82**	8 (11.6%)	67 (90.5%)
9	**.36**	172	**0.60**	09 (11.1%)	61 (67.0%)	**.37**	145	**0.70**	04 (05.6%)	55 (75.3%)
10	**.50**	169	**0.59**	04 (04.7%)	54 (65.1%)	**.36**	143	**0.71**	8 (11.9%)	59 (77.6%)
11	**.28**	175	**0.53**	23 (28.4%)	70 (74.5%)	**.38**	151	**0.62**	03 (04.1%)	50 (64.9%)
12	**.43**	159	**0.76**	27 (41.5%)	88 (93.6%)	**.47**	147	**0.80**	02 (02.9%)	61 (77.2%)
13	-.31	179	-0.27	61 (70.9%)	36 (38.7%)	**.43**	155	**0.77**	23 (32.9%)	83 (97.6%)
14	.12	182	0.02	00 (00.0%)	02 (02.2%)	.14	141	0.22	01 (01.3%)	17 (25.8%)
15	.22	161	0.27	06 (07.6%)	28 (34.1%)	**.45**	144	**0.76**	02 (02.9%)	57 (76.0%)
16	**.42**	170	**0.36**	01 (01.1%)	32 (38.6%)	**.43**	150	**0.78**	20 (29.9%)	79 (95.2%)
17	**.35**	176	**0.33**	01 (01.1%)	30 (34.9%)	**.28**	157	**0.35**	06 (07.6%)	34 (43.6%)
18	**.50**	148	**0.70**	02 (03.0%)	54 (65.9%)	**.29**	157	**0.42**	02 (02.6%)	35 (43.8%)
19	**.39**	174	**0.50**	04 (04.8%)	48 (52.7%)	.17	164	0.08	00 (00.0%)	07 (08.6%)
20	.18	178	0.14	05 (05.7%)	18 (20.0%)	**.28**	160	**0.30**	61 (81.3%)	85 (100%)
21	**.39**	152	**0.60**	06 (08.8%)	52 (61.9%)					
22	.28	182	0.08	00 (00.0%)	08 (08.8%)					
23	**.32**	167	**0.25**	00 (00.0%)	21 (25.9%)					
24	**.41**	166	**0.34**	02 (02.4%)	31 (38.3%)					
25	.22	183	0.16	01 (01.1%)	16 (17.2%)					
26	**.30**	146	**0.61**	24 (37.5%)	69 (84.1%)					
27	.27	175	0.13	00 (00.0%)	12 (13.8%)					
28	.03	173	0.23	22 (27.5%)	42 (45.2%)					
29	.15	179	0.13	00 (00.0%)	12 (13.2%)					
30	**.29**	149	**0.57**	16 (25.8%)	59 (67.8%)					
31	**.42**	163	**0.74**	23 (31.9%)	84 (92.3%)					
32	**.44**	146	**0.52**	00 (00.0%)	38 (48.1%)					
33	.10	165	0.17	04 (05.0%)	18 (21.2%)					
34	.18	174	0.18	00 (00.0%)	16 (18.8%)					
35	**.29**	158	**0.32**	05 (06.7%)	31 (37.3%)					
36	.32	169	0.18	00 (00.0%)	16 (19.3%)					
37	**.43**	168	**0.60**	03 (03.6%)	54 (63.5%)					
38	**.42**	163	**0.65**	14 (17.5%)	67 (80.7%)					
39	**.41**	165	**0.58**	09 (11.5%)	57 (65.5%)					
40	**.35**	167	**0.61**	35 (47.3%)	86 (92.5%)					

Quartile groupings and *d* are discrimination indices.

Section 5.2.1.3: Item Difficulty Indices

After determining item discrimination power characteristics item difficulty levels for retained intelligence items were calculated. As was noted earlier, the most common index of item difficulty is the proportion of respondents that answered the question correctly (Kline, 2000). Table 5.12 provides information on the proportion of respondents that answered fluid and crystallised intelligence items correctly. For continuing retention fluid and crystallised intelligence items required a proportion of correct responses ranging from .1 to .9. Table 5.12 also details the item difficulty index of logit *p,* which is the natural logarithm of the correct to incorrect ratio (Brown & White, 2009). This index is not used as a criterion for item retention, but is used in subsequent hypothesis testing.

Table 5.12 indicates that the 15 crystallised intelligence items having met the preceding selection criteria for item discrimination also fell within the .1 to .9 proportion correct range (1, 3, 6, 14-16, 18-20, 22, 24-27, 29). The proportion correct values of these items ranged from .22 to .85.

Table 5.12 also suggests the 17 fluid intelligence items having met the preceding selection criteria for item discrimination also fell within the .1 to .9 proportion correct range (1, 2, 4, 8, 9, 10-12, 16, 19-26). The proportion correct values of these items ranged from .19 to .85.

Table 5.12

Difficulty Indices for Crystallised and Fluid Ability Items

Crystallised Items				Fluid Items			
Items	N	Prop. C	Log C/E	Items	N	Prop. C	Log C/E
1	315	**.47**	-0.03	1	315	**.85**	0.76
3	314	**.29**	-0.38	2	308	**.19**	-0.61
6	315	**.85**	0.75	4	314	**.65**	0.27
14	316	**.73**	0.43	8	310	**.52**	0.04
15	314	**.84**	0.73	9	287	**.19**	-0.61
16	315	**.22**	-0.53	10	314	**.57**	0.12
18	313	**.38**	-0.20	11	299	**.56**	0.12
19	317	**.43**	-0.11	12	316	**.71**	0.40
20	313	**.51**	0.01	16	311	**.58**	0.15
22	316	**.62**	0.22	19	312	**.64**	0.25
24	315	**.79**	0.57	20	314	**.72**	0.41
25	316	**.65**	0.27	21	310	**.64**	0.26
26	316	**.44**	-0.08	22	314	**.82**	0.66
27	312	**.43**	-0.11	23	314	**.71**	0.39
29	314	**.80**	0.62	24	303	**.51**	0.02
				25	311	**.65**	0.28
				26	309	**.73**	0.44

Note. Listwise deletion of missing data. "Prop. C" = proportion correctly answering item. Proportion correctly answering the item and the Log of correct divided by incorrect responses serve as the discrimination indices.

Section 5.2.1.4: Item Distractor Analysis

Distractor analysis of ability items was undertaken after identifying which items met the discrimination and difficulty retention criteria. In multiple-choice ability items, a distractor is an option deliberately representing an incorrect answer, in order to attract a less confident respondent's attention "away" from the correct answer.

Table 5.13 shows the distractor analysis for crystallised intelligence items. The analysis indicates that for all items more respondents from the upper one-quarter group chose the correct response relative to the lower one-quarter group. Table 5.13 also shows that more upper one-quarter respondents chose the correct response rather than distractors for all items. Despite this result a number of items that were acceptable in terms of difficulty and discrimination indices may benefit from some improvements or adjustment for their distractors. Three items had an implausible distractor not chosen by any lower quartile respondents (14, 19, and 29). Items 14 and 19 also appear to have a disproportionately attractive distractor. These items may benefit from distractor adjustments, such as rewording or replacing the distractor in question.

Table 5.14 shows the distractor analysis for fluid intelligence items. The analysis indicates that for all items more respondents from the group under the first quartile versus the group over the third quartile chose the correct response relative to the lower quartile group. A number of items could still potentially benefit from some improvements or adjustment for their distractors. Item two had an implausible distractor not chosen by any lower quartile respondents, while items one and 22 may have some disproportionately attractive distractors. These items may benefit from distractor adjustments, amendments, or outright replacements.

Table 5.13

Crystallised Intelligence Item Distractor versus Correct Response-Option Analysis for Upper and Lower Quartiles

		Crystallised Item Response Options				
	Quartile	1	2	3	4	5
Item 1	U	9	0	0	6	**60**
	L	20	4	4	35	**13**
Item 3	U	7	2	0	**46**	20
	L	24	4	2	**13**	32
Item 6	U	**71**	0	1	3	0
	L	**44**	3	2	24	2
Item 14	U	**68**	0	7	0	0
	L	**38**	6	29	0	4
Item 15	U	0	**74**	0	2	0
	L	2	**46**	10	12	7
Item 16	U	0	33	**36**	0	5
	L	1	23	**5**	7	41
Item 18	U	4	1	13	1	**54**
	L	17	4	43	8	**7**
Item 19	U	0	1	16	**58**	0
	L	5	14	52	**8**	0
Item 20	U	7	5	2	**62**	1
	L	34	11	7	**22**	2
Item 22	U	5	**67**	1	0	2
	L	17	**24**	21	1	14
Item 24	U	0	1	**71**	1	1
	L	2	10	**41**	8	17
Item 25	U	4	**65**	5	1	0
	L	9	**29**	17	12	12
Item 26	U	17	0	**58**	0	0
	L	33	18	**15**	7	4
Item 27	U	2	**58**	2	12	0
	L	11	**17**	23	17	7
Item 29	U	2	0	0	0	**73**
	L	5	0	1	29	**42**

Note. "U" = Upper Quartile, "L" = Lower Quartile. No cases deleted. Correct response-option is emboldened and number represents number of individuals selecting option.

Table 5.14

Fluid Intelligence Item Distractor versus Correct Response-Option Analysis for Upper and Lower Quartiles

		Fluid Item Response Options				
Items	Quartile	1	2	3	4	5
Item 1	U	**51**	14	18	1	**3**
	L	**29**	4	36	3	**4**
Item 2	U	0	**86**	2	**0**	0
	L	2	**53**	18	**2**	0
Item 4	U	0	0	0	**88**	1
	L	1	11	14	**37**	12
Item 8	U	0	2	**84**	0	2
	L	12	20	**28**	7	7
Item 9	U	2	1	**65**	15	2
	L	13	18	**34**	2	8
Item 10	U	20	6	4	18	**30**
	L	22	16	12	17	**6**
Item 11	U	9	**56**	**5**	11	2
	L	11	**26**	**6**	24	7
Item 12	U	1	0	0	**83**	4
	L	12	8	12	**12**	31
Item 16	U	1	3	**81**	1	2
	L	30	10	**22**	8	3
Item 19	U	0	0	0	**87**	1
	L	1	2	2	**65**	5
Item 20	U	2	**2**	6	7	**68**
	L	3	**7**	20	21	**23**
Item 21	U	0	8	**57**	11	10
	L	2	1	**38**	5	27
Item 22	U	**38**	19	3	27	0
	L	**33**	7	4	28	3
Item 23	U	2	1	21	16	**48**
	L	2	1	17	23	**32**
Item 24	U	1	4	4	**73**	5
	L	4	7	14	**42**	8
Item 25	U	0	0	0	**85**	3
	L	2	4	2	**33**	35
Item 26	U	17	**40**	11	2	12
	L	25	**20**	12	7	8

Note. "U" = Upper Quartile, "L" = Lower Quartile. No cases deleted. Correct response-option is emboldened and number represents number of individuals selecting option.

Section 5.2.2: Scale-level Analysis

Once CTT item analysis was complete the analysis moved to more scale-level considerations. The first of these involved the use of factor analysis in order to establish factor structure – a tentative assessment of construct validity. The second involved the use of Cronbach's α to determine how internally consistent scales were. The third involved looking at inter-scale correlations as a tentative assessment of discriminant validity. The decision to undertake factor analysis before estimating reliability and discriminant validity was based upon factor loadings serving as the final hurdle for personality item retention. Item response theory derived item parameters served as the final hurdle for fluid and crystallised intelligence retention.

Section 5.2.2.1: Validity Check

In line with Nunnally's (1978) recommendation only items that met item analysis retention criteria were factored. Before conducting factor analysis it is important to confirm the factorability of the data in question. The first step in this process involved confirming the approximate normality of distributions for items. The second step involved producing a correlation matrix which was then examined for suspiciously high correlations likely to reduce the factorability of the data. No such relationships were observed for either personality or ability items. The KMO measure of sampling adequacy was .73 for personality items and .74 for personality items. The chi-square assessments of Bartlett's test of sphericity were significant for both ability and personality items. These results suggest the data was suitable for factor analysis. The anti-image matrix was also examined for any results below the required .5. No ability items were found to have anti-image matrix results below .5. Conscientiousness item 28 and Integrity item 2 were both found to have anti-image values below .5 (.36 and .39 respectively). These items were excluded from subsequent analysis. Communalities were also checked to ensure all variables shared more than 20% of their variance in common with other variables. This is important in assisting subsequent factor identification and labelling (Child, 2006).

Once the factorability of data had been confirmed the number of factors to extract was determined via *Principal Components Analysis* (PCA). Principal Components Analysis looks to maximise the amount of variance explained in the items from a purely descriptive perspective and is recommended for determining the number of factors worthy for inclusion in subsequent analysis (Child, 2006; Tabachnick & Fidell, 2001).

Section 5.2.2.1.1: Principal Component Analysis of Cognitive Ability Items

The 32 crystallised and fluid intelligence items to have met the discrimination and difficulty criteria were included in a PCA. The PCA suggested that 12 components had eigenvalues above one. It also suggests the first principal component could account for 15.37% of variance. The finding of a single large general factor in PCA and principal factor analysis is an artefact of the algebraic procedures and should not be given excessive consideration prior to rotations (Kline, 1994). It is only after factor rotations when variance from the general factor is redistributed that factor structures are likely to become more accurately interpretable (Kline, 1994). The other 11 components exceeding an eigenvalue 1 individually accounted for between 6.56 and 3.13% of variance. These 12 components cumulatively account for 60.03% for variance.

Figure 5.1's scree plot shows the eigenvalue for each of the 12 principal components. The "elbow" suggests a substantial difference between the amount of variance explained by the first principal component and subsequent components. Scree-based criterion suggests that only the first two factors which are above the "elbow" are considered salient factors (Manly, 2005).

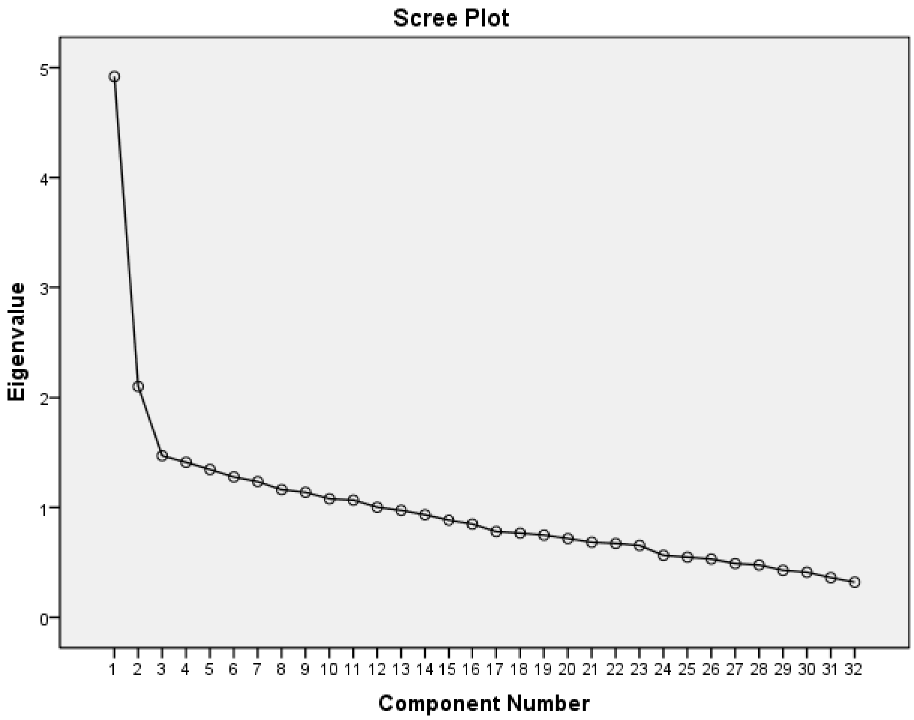


Figure 5.1. Fluid and Crystallised Intelligence Items Scree Plot.

Section 5.2.2.1.2: Common Analysis of Extracted Cognitive Ability Factors

After determining the number of salient factors to extract via PCA it becomes possible to conduct a factor analysis to detect theoretically relevant underlying structures within the data. All factor analyses and factor rotations used both *Principal Axis Factoring* (PAF) and *Maximum Likelihood Factor Extraction* (ML) for comparative purposes. Common analysis of the 12 components with eigenvalues of > 1 for the two salient factors identified via the scree plot was also undertaken before PAF and ML. This was intended to serve as an additional check on whether it was appropriate to limit the subsequent analyses to two factors. This analysis showed that the two factors above the scree elbow were the only factors to have three or more unique factor loadings, which was noted earlier as the minimum required number to define a factor (Child, 2006).

Factor matrices for PAF and ML both showed 21 unique factor loadings on the first factor and three on the second factor. Four of the unique factor loadings on the first factor relate to crystallised intelligence items. Three of these related to "odd man out" items. The fourth asked what an "eddy" is. The remaining 17 factor loadings for the first factor covered all of the fluid intelligence items. All three factor loadings for the second factor came from the crystallised intelligence items. Two of these were antonym identification items. The third asked respondents what a "curmudgeon" is. Principal axis factoring and ML direct factor analysis produced very consistent results. The next step undertaken was factor rotation (Child, 2006).

Section 5.2.2.1.3: Factor Rotation of Extracted Cognitive Ability Factors

Both PAF and ML oblique rotation produced very similar results. Considerable similarity between the pattern and structure matrices was also observed. The component correlation matrices showed small relationships between the two factors extracted. Supporting Tables are reported in Appendix C. On the basis of these findings orthogonal rotation was justified and no second-order factor rotations were recommended (Child, 2006).

Table 5.15 reports the orthogonal rotated matrices. There appears little difference between these rotations and those reported in Appendix C. The most important items for the identification of factor one are fluid items 19, 22, 23, and 26. Items 27 and 24 have the strongest loading on factor two, followed by items 22 and 18. Crystallised intelligence item 19 and fluid intelligence item 20 failed to demonstrate unique factor loadings. For this reason crystallised intelligence items 19 and 20 were excluded from subsequent analysis (Gorsuch, 1983; Thompson, 2004).

Table 5.15

Cognitive Ability Rotated Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

	PAF Factor Rotation		ML Factor Rotation	
Items	1	2	1	2
Crystallised				
1		.37		.36
3				.25
6		.29		.28
14	.29			
15		.25		.25
16				
18		.38		.40
19	.40	.37	.39	.39
20		.28		.27
22		.34		.33
24		.44		.42
25		.28		.28
26		.25		.27
27		.45		.46
29				
Fluid				
1	.38		.37	
2	.31		.31	
4	.44		.44	
8	.29		.28	
9	.27		.27	
10	.45		.45	
11	.42		.42	
12	.44		.44	
16	.30		.29	
19	.63		.63	
20	.32	.32	.33	.32
21	.36		.36	
22	.58		.58	
23	.49		.49	
24	.44		.43	
25	.38		.38	
26	.51		.51	

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 3 iterations. Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 3 iterations.

Section 5.2.2.1.4: Principal Component Analysis of Personality Items

The 112 personality items that met the discrimination and difficulty criteria were included in a PCA. The numbers of items included for each scale were: conscientiousness (30), integrity (25), neuroticism (42), and dissimulation (15). As previously mentioned, dissimulation is an administration scale, not a personality trait. Thirty-six components had eigenvalues above one. The first component accounts for

13.93% of variance. The other 35 components exceeding eigenvalue 1 accounted for between 5.73 and 0.89% of variance. These 36 components cumulatively accounted for 70.36% for variance.

Figure 5.2's scree plot shows the eigenvalue for each of the principal components. The "elbow" suggests a substantial difference between the amount of variance explained by the first principal component and subsequent components. Five components in Figure 5.2 met the earlier discussed above the "elbow" criterion (Manly, 2005).

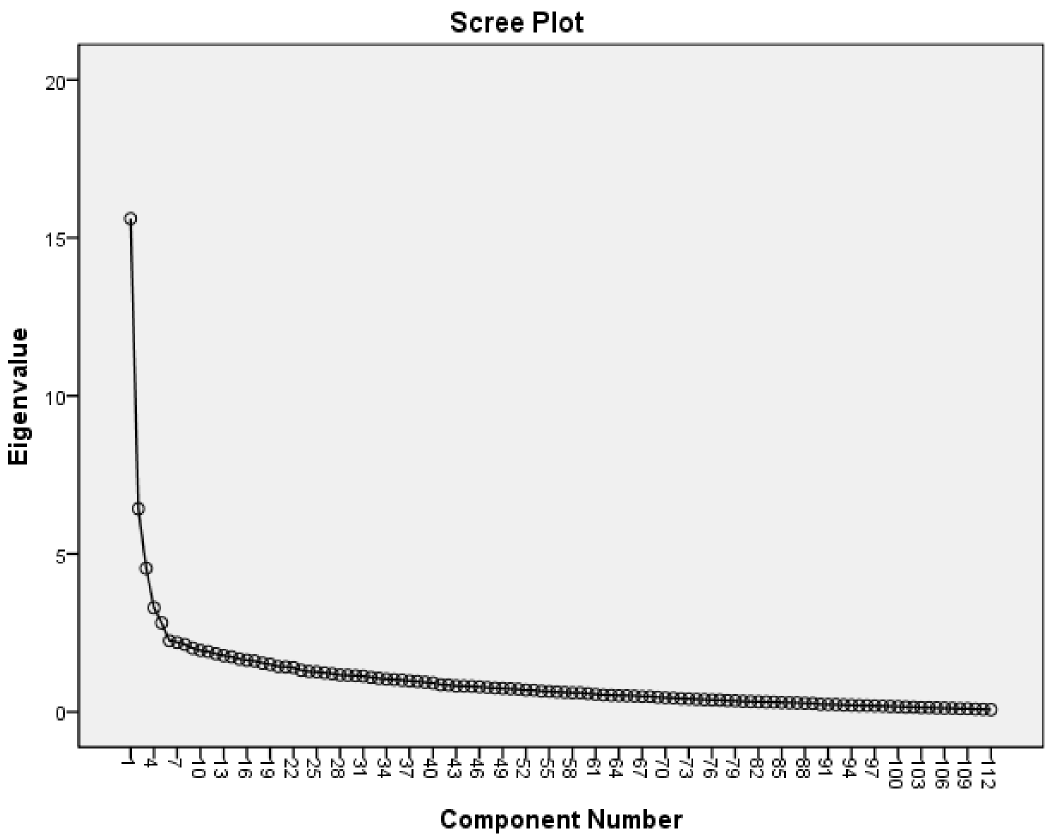


Figure 5.2. Personality Item Scree Plot.

Section 5.2.2.1.5: Common Analysis of Extracted Personality Factors

As with the ability items, after determining the number of salient personality factors to extract via PCA common analysis of extracted factors was undertaken. Once again, results were compared across PAF and ML. As with the ability items, common analysis of all 36 components with eigenvalues <

1was performed before PAF and ML. This showed that no additional factors met the requirement for defining a factor of three or more unique factor loadings (Child, 2006).

Factor matrices for PAF and ML both showed very similar factor loadings across all five factors. The first factor has 42 (PAF) and 40 (ML) unique factor loadings above .3. The second factor had 10 unique factor loadings according to both PAF and ML extraction. The third factor had 6 (PAF) and 8 (ML) unique loadings. The fourth factor had 4 (PAF) and 3 (ML) unique loadings. Neither model had any unique factor loadings for the fifth factor. Factor loadings for the first factor span items from all scales. Factor loadings from conscientiousness and dissimulation items were all in the negative direction. Factor loadings from neuroticism and integrity were all in the positive direction. Factor loadings from the second factor span conscientiousness, neuroticism, and dissimulation items. These factor loadings were all in the positive direction. Factor loadings for factor three were all in the positive direction and restricted to integrity scale items. Factor loadings for the fourth factor were all in the positive direction and restricted to dissimulation items. The fifth factor had mixed positive and negative loadings restricted to neuroticism scale items.

Section 5.2.2.1.6: Factor Rotation of Extracted Personality Factors

As with ability items, further factor analysis with factor rotation was undertaken to facilitate factor interpretation and item retention. Both PAF and ML oblique rotation produced very similar results. There was also considerable similarity between the pattern and structure matrices for personality factors. As with those of the ability items, the component correlation matrices showed small relationships between the two factors extracted. This is consistent across both PAF and ML methodology. On this basis orthogonal rotation was justified (Thompson, 2004). Again, it also means no second-order factor rotations are recommended (Child, 2006). Supporting Tables are reported within Appendix C.

Table 5.16 contains the orthogonal rotated matrices. There appeared little overall difference between these and oblique rotations. The one major difference is that the items loading on factor one from the oblique rotations appears to have replaced the content of what was factor four, while the previous content of factor four has moved to factor five, and the previous content of factor five has moved to factor one (see Appendix C). Factor four in Table 5.16 is strongly defined by neuroticism item factor loading that are characterised by mercurial emotionality (e.g., item 23 “I am quick to anger). While the content of factor one is defined by anxiety (e.g., item 5 “I often worry about things”). For subsequent analyses factors one and four below will be referred to as neuroticism angst factor and neuroticism mercurial factor. Factor five is most strongly defined by dissimulation loadings.

Table 5.16

Personality Rotated Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

	PAF Rotated Matrix					ML Rotated Matrix				
Items	1	2	3	4	5	1	2	3	4	5
Con										
1		.39					.36			
3		.69					.70			
4		.51					.51			
5										
8	-.38					-.38				
10		.42					.42			
11		.40					.39			
12		.31								
15		.38					.37			
16		.61					.60			
19		.40					.39			
23		.52					.53			
24										
26		.30								
27		.54					.55			
28										
29		.45					.47			
30		.42					.42			
31		.60					.61			
32										
34		.32					.31			
35		.45					.45			
36					.32					.31
41										
42		.53					.54			
45		.44					.42			
46										
48		.35					.34			
50		.55					.56			
N										
2	.44					.44				
3								.30		
5	.60					.60				
6				.38					.39	
7	.43					.43				
8	.36					.36				

9				.56					.59	
10									.31	
11				.46					.47	
12	.39			.33		.39			.34	
13	.49					.48				
14			.38	.38				.39	.39	
15	.45					.45				
16				.47					.44	
17	.42					.43				
18	.51					.50				
19				.38					.37	-.30
20	.38			.42		.37			.43	
22		-.33					-.33			
23				.52					.50	
24	.58					.57				
25	.46					.46				
26			.38					.39		
28	.36					.37				
30	.44					.44				
31										
32	.50					.51				
33	.35			.44		.35			.44	
34	.50					.50				
35	.38			.47		.37			.50	
36			.34	.36				.35	.36	
38										
39	.57					.58				
40	.50					.49				
41				.45					.44	
42	.32					.32				
43	.63					.63				
44	.54					.54				
46				.37					.37	
47	.48					.48				
49			.40					.40		
50	.33					.32				
In										
1			.53					.52		
3			.52					.51		
4			.31					.31		
7			.48					.48		
8			.36					.36		
9			.44					.44		
10			.55					.56		
11										
12			.44					.43		
16			.41					.41		
17			.40					.40		
18			.54					.53		
19			.40					.40		
21			.47					.47		
24			.38					.37		
26			.30					.30		
30			.32					.32		
31			.40					.40		
32			.42					.41		
35										
37			.51					.52		
38			.45					.45		

39			.40					.38		
40										
D										
1					.45					.44
3					.30					
5					.36					.35
6										
7					.40					.38
8					.48					.48
9					.37					.36
10					.47					.48
11					.38					.37
12		.37			.47		.35			.48
13		.59					.59			
15	-.34				.35	-.34				.36
16		.36			.34		.36			.35
17					.31					.30
18					.45					.46

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 6 iterations. Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 7 iterations.

Based upon content analysis and strength and uniqueness of factor loadings 20 conscientiousness (items 1, 3, 4, 10, 11, 12, 15, 16, 19, 23, 27, 29, 30, 31, 34, 35, 42, 45, 48, and 50), 20 neuroticism (items 5, 6, 9, 11, 13, 15, 16, 18, 23, 24, 25, 30, 32, 34, 39, 40, 41, 43, 44, and 47), 20 integrity (items 1, 3, 4, 7, 8, 9, 10, 12, 16, 17, 18, 19, 21, 24, 30, 31, 32, 37, 38, and 39), and 10 dissimulation (items 1, 3, 5, 7, 8, 9, 10, 11, 17, and 18) items were selected for inclusion in the refined assessments (see Appendix A). The outcomes of these exploratory factor analyses provide tentative support for construct validity of assessed scales, yet tentative support for the reliability of these refined scales was still required.

Section 5.2.2.2: Reliability Check

Table 5.17 provides corrected Cronbach's α values for the internal consistency of each scale. The scale for crystallised Intelligence is the only one that fails to reach the .7 recommended standard of reliability (Eatwell, 1997; *cf*. Lance, Butts, & Michels, 2006).

Table 5.17

Cronbach's Alpha for Retained items

Scale	Alpha	Items	N
Crystallised Intelligence	.66	14	289
Fluid Intelligence	.78	16	212
Integrity	.88	20	281
Conscientiousness	.82	20	282
Neuroticism	.89	20	280
Dissimulation	.75	10	305

Reliability of each scale was also estimated separately according to gender. Cronbach's α for male versus female respondents are reported in Table 5.18. The personality scale α across these groups were very similar. The ability α suggest slight to moderate differences. The crystallised intelligence α suggest that items may be slightly more consistent in their measurement of crystallised intelligence for males than females. The fluid intelligence α suggest that items are likely to be more consistent in their measurement of fluid intelligence for females than males.

Table 5.18

Cronbach's Alpha Comparison between Males and Females

Scale	Items	Male α	N	Female α	N
Crystallised Intelligence	14	.68	108	.60	177
Fluid Intelligence	16	.67	73	.80	140
Integrity	20	.90	110	.86	177
Conscientiousness	20	.88	106	.88	179
Neuroticism	20	.89	111	.89	186
Dissimulation	10	.81	115	.75	190

Section 5.2.3: IRT Calibration of Retained Ability Items

Fluid and crystallised intelligence items were also subject to analysis guided by the assumptions of item response theory (Rasch, 1980). This was done to determine item weights (Lord, 1980) and extend CTT reliability and validity (Ackerman, 1992; Embretson & Reise, 2000).

An evaluation of model-data fit was undertaken using chi-squares to test for both global fit and local independence (Embretson & Reise, 2000). Fit analysis involves comparing expected and observed frequencies of item category responses for various levels of the IRT score (Bjorner, Smith, Stone, & Sun, 2007). The global fit for crystallised and fluid intelligence items to the 3PL model was good. All items for both fluid and crystallised scales were below the 3.0 chi-square value recommended (Drasgow, Levine, Tsien, Williams, & Mead, 1995). Furthermore, no significant global-level relationships were reported (See Appendix D for supporting Table).

Tests of local item independence explore the assumption that the probability of a respondent giving a correct response for any item is a function of a respondent's ability level, not the respondent's answers to other items (Chernyshenko et al., 2008). The tests of local independence undertaken in this investigation suggested potential issues with two crystallised intelligence items (18 and 26) and ten fluid intelligence items (4, 9, 11, 12, 16, 19, 21, 23, 24, and 26). Item mis-fit is often due to poor item quality concerning distractor issues or construct relevance (Barker, 2001). The two crystallised intelligence items with apparent local dependence were previously identified as candidates for distractor revision. Item 26 also failed to load above .3 on the factor of crystallised intelligence identified in previous factor analysis. Four of the fluid intelligence items flagged also had been recommended for distractor revision (12 and 16) and/or failed to load consistently and importantly during factor analysis (9 and 16). The cause of local dependence for fluid items 4, 11, 19, 21, 23, 24, and 26 are less apparent.

Tests of local independence are often prone to false positives (Ip, 2000, 2001). Furthermore, according to Hambleton, Swaminathan, and Rogers (1991) data should not be removed on the basis of simply failing a test of local independence. Items should instead be removed for construct-relevant reasons. No such reasons were identified upon examination of the item characteristic curves, statistic parameters, and information function curves for these items.

After testing the model-data fit, 3PL item parameters were calculated and evaluated. The strength of item discrimination (*a*) parameters is reported according to the following approximate guide: low ($a = 0.50$), medium ($a = 1.0$), and high ($a = 2.0$) (Chernyshenko et al., 2008). As can be seen in Table 5.19, *a*-parameters for crystallised ability items ranged from 0.737 to 3.400 and 0.673 to 2.023 for fluid intelligence.

Table 5.19

Crystallised Intelligence Parameter Estimates

Item	Slope (a)	s.e.	Threshold (b)	s.e.	Asymptote (c)	s.e.
1	1.14	0.20	0.53	0.14	0.13	0.01
3	3.40	0.70	1.37	0.10	0.18	0.00
6	1.32	0.23	-1.43	0.23	0.15	0.05
14	0.74	0.16	-1.07	0.33	0.16	0.05
15	1.10	0.20	-1.52	0.29	0.16	0.06
16	1.59	0.38	1.80	0.23	0.12	0.00
18	1.27	0.22	0.88	0.14	0.11	0.00
20	0.99	0.20	0.64	0.18	0.19	0.01
22	1.17	0.19	-0.19	0.14	0.15	0.02
24	0.90	0.18	-1.33	0.31	0.16	0.06
25	0.74	0.16	-0.49	0.26	0.15	0.04
26	1.24	0.24	0.82	0.15	0.18	0.01
27	1.21	0.22	0.82	0.15	0.14	0.01
29	0.73	0.16	-1.72	0.44	0.16	0.06

Table 5.19 and Table 5.20 also detail item difficulty (*b*) and pseudo-guessing (*c*) parameters. The *c*-parameter reflects the lower asymptote of an item characteristic curve (ICC). For items having four or five response options the *c*-parameter generally ranges from .1 to .3 (Chernyshenko et al., 2008). The *c*-parameters for crystallised intelligence items range from .11 to .19. Furthermore, the *b*-parameters in Table 5.19 suggest a relatively good spread of difficulty amongst crystallised intelligence items. Specifically, crystallised items are shown to range in difficulty between -1.72 and 1.80.

The item characteristic curve comparisons for the most and least discriminating crystallised intelligence items are presented in Figure 5.3. Figure 5.4 provides comparative ICCs for the most and least difficult crystallised intelligence items. (Item characteristic curves for all ability items can be found in Appendix D. Item information functions for each item are available on request.)

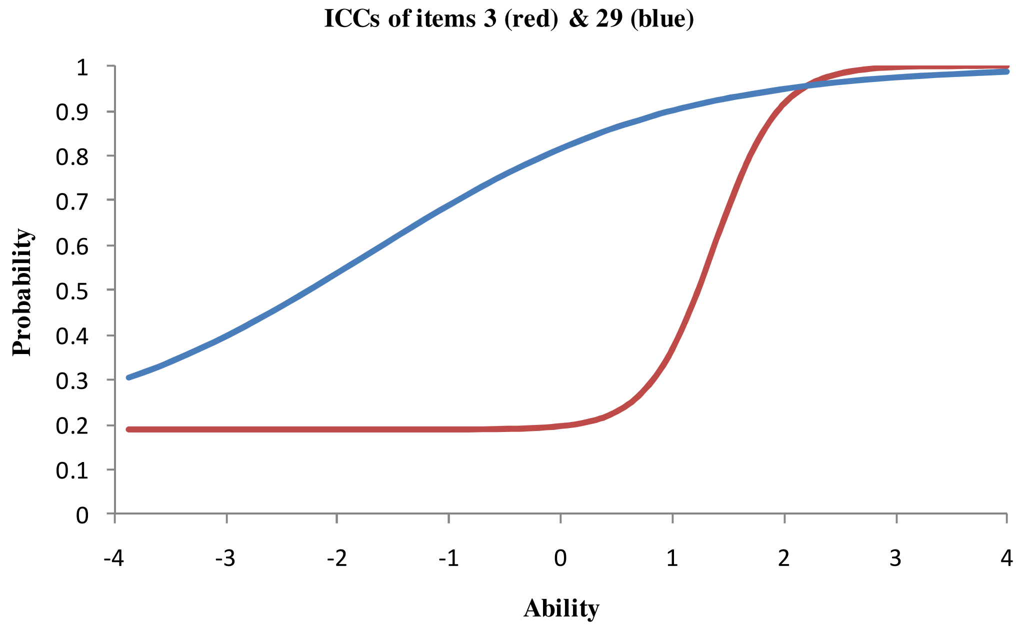


Figure 5.3. Item Characteristic Curves for Most and Least Discriminating Crystallised Ability Items (red and blue respectively).

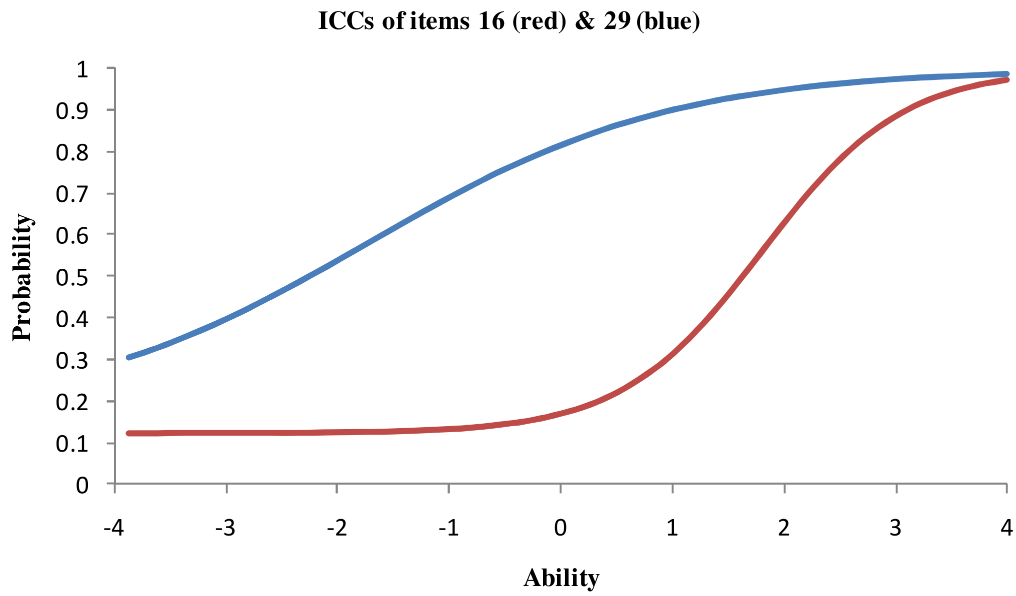


Figure 5.4. Item Characteristic Curves for Most and Least Difficult Crystallised Ability Items (red and blue respectively).

The *c*-parameters for fluid intelligence items in Table 5.20 range from .05 through to .18. Furthermore, the *b*-parameters in Table 5.20 suggest a relatively good spread of difficulty amongst fluid intelligence items. Specifically, fluid intelligence items have difficulty levels ranging from -1.74 to 2.49.

Table 5.20

Fluid Intelligence Parameter Estimates

Item	Slope (a)	s.e.	Threshold (b)	s.e.	Asymptote (c)	s.e.
1	0.91	0.18	-1.74	0.39	0.16	0.07
2	1.66	0.31	1.70	0.16	0.05	0.00
4	1.07	0.18	-0.18	0.17	0.15	0.02
8	0.97	0.18	0.58	0.16	0.15	0.01
9	1.07	0.29	2.49	0.42	0.07	0.00
10	1.07	0.17	0.16	0.14	0.12	0.02
11	0.74	0.16	0.46	0.21	0.14	0.02
12	1.14	0.19	-0.53	0.18	0.16	0.03
16	0.67	0.15	0.21	0.24	0.16	0.03
19	2.02	0.28	-0.05	0.09	0.12	0.00
21	1.06	0.19	-0.02	0.16	0.18	0.02
22	1.70	0.26	-0.92	0.15	0.15	0.03
23	1.32	0.20	-0.48	0.15	0.14	0.02
24	0.94	0.18	0.67	0.16	0.13	0.01
25	0.79	0.15	-0.30	0.23	0.14	0.03
26	1.09	0.18	-0.62	0.19	0.14	0.03

The item characteristic curve comparisons for the most and least discriminating fluid ability items are presented in Figure 5.5. Figure 5.6 provides comparative ICCs for the most and least difficult fluid intelligence items.

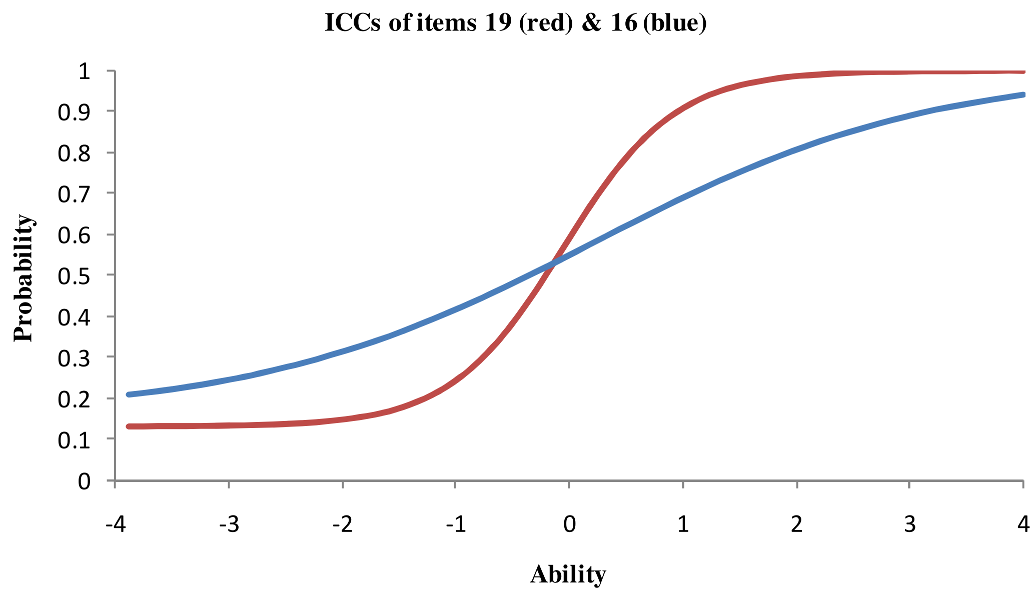


Figure 5.5. Item Characteristic Curves for Most and Least Discriminating Fluid Ability Items (red and blue respectively).

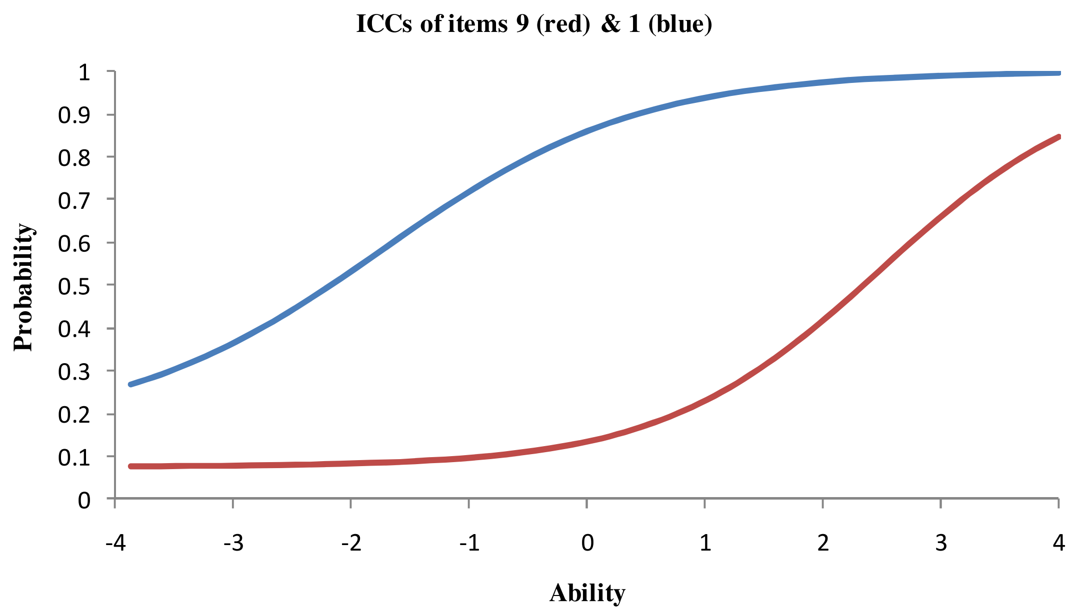


Figure 5.6. Item Characteristic Curves for Most and Least Difficult Fluid Intelligence Items (red and blue respectively).

Once item parameters were determined the precision of the assessments and items were examined via item information functions (IIF). Item information functions tend to look symmetrical

(Hambleton et al., 1991). Highly discriminating items will have narrow, tall information functions. These items contribute greatly but within a narrow range of ability. Less discriminating items provide less information but cover a wider range of ability. According to Baker (2001) when three-parameter models are used values of the guessing parameter *c* higher than zero lower the amount of test information at the low-ability levels. Furthermore, large values of *c* reduce the general level of the amount of test information. Overall IIFs for crystallised and fluid intelligence items are provided in Figure 5.7 and Figure 5.8. These information functions should be compared to the target information functions of these assessments (Hambleton & Jones, 1994), which is the average range of ability.

The location of the peak in Figure 5.7 suggests the crystallised intelligence assessment will be most accurate for those with an above-average level of ability, which is greater than the target information function of this assessment. However, as *a* values > 1.7 indicate a high general level of ability information (Baker, 2001), the slant of the left-hand slope away from this peak suggests that the assessment will also provide precise information on those in the slightly below-average range. Moreover, it is not until the well below or above average range is reached that the assessment is reduced to a low general level of information (*a* < 1.0).

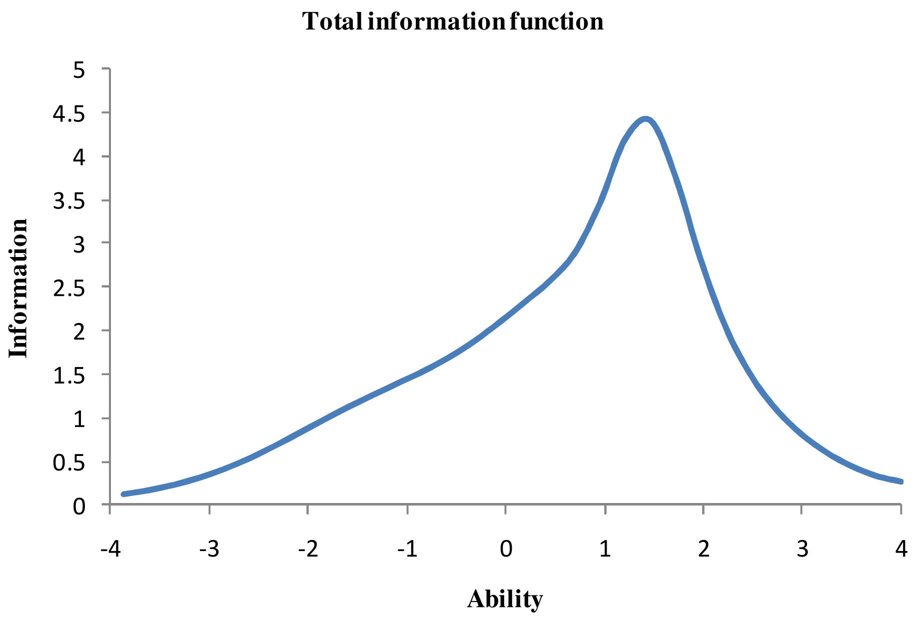


Figure 5.7. Crystallised Ability Total Information Function.

Consistent with the target information function of this assessment, the location of the peak in Figure 5.8 suggests the fluid intelligence assessment will be most accurate for those with an average level of ability. The slant of slopes in Figure 5.8 indicates that the assessment will provide precise information on those ranging in ability from below average to well above average. It is not until the well below or superior range is reached that the assessment is reduced to a low general level of information ($a < 1.0$).

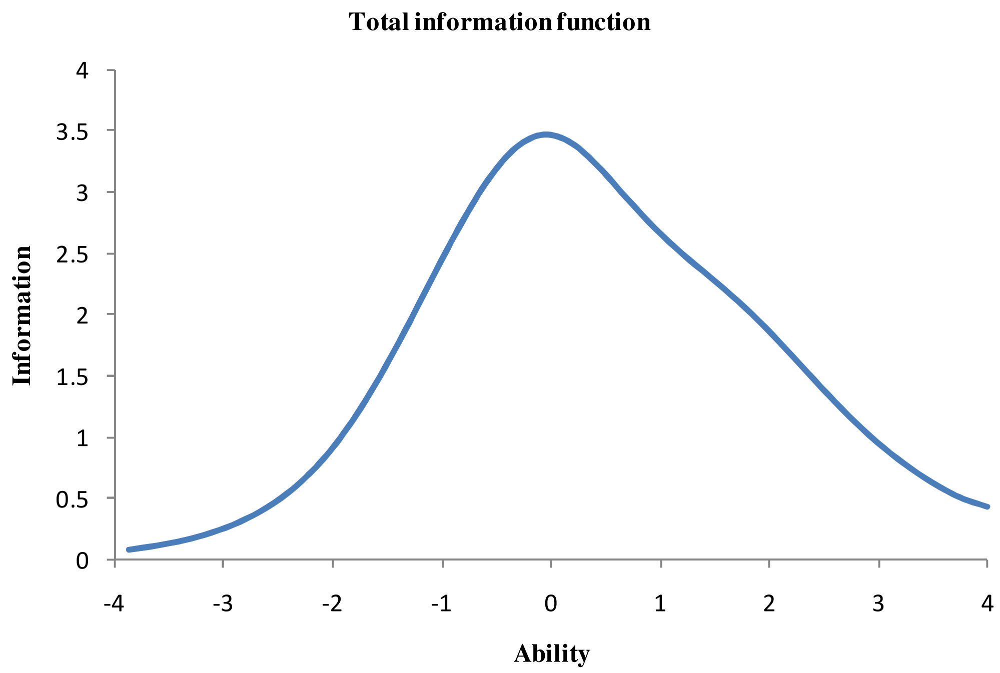


Figure 5.8. Fluid Ability Total Information Function.

Section 5.2.4: Refined Scale Characteristics

Characteristics of refined fluid and crystallised intelligence, conscientiousness, integrity, neuroticism, and dissimulation scales are detailed in this section. Table 5.21 details the descriptive statistics for the finalised intelligence and personality scales. Intelligence scale scores were weighted on the basis of IRT modelled discrimination and personality scale scores were refined on the basis of factor loadings.

Table 5.21

Scale score means and standard deviations

Scale	Item N	Min.	Max.	Mean	Std. Deviation
Crystallised Intelligence	14	0.73	17.60	9.04	3.49
Fluid Intelligence	16	0.00	18.29	10.72	3.89
General Intelligence	30	5.85	33.35	19.76	5.98
Integrity	20	20	88	51.53	11.84
Conscientious	20	38	94	69.09	11.26
Neuroticism	20	23	88	56.35	12.17
Dissimulation	10	10	45	26.38	5.90

Note. N = 317.

Table 5.22 details the relationships between final weighted and original raw scale scores for fluid and crystallised intelligence. The relationships between final and raw crystallised intelligence scale scores are significant and strong ($r = .87, p < .001$). Fluid intelligence raw and finalised scale scores are also strongly and significantly correlated ($r = .93, p < .001$).

Table 5.22

Fluid and Crystallised Intelligence Refined versus Raw Scale Scores

	Crystallised Weighted Total	Fluid Weighted Total
Crystallised Intelligence Raw Total	**.87***	**.37***
Fluid Intelligence Raw Total	**.34***	**.93***

Note. Correlations marked * are significant at p < .007, N=317, (Listwise deletion of missing data).

Details of the relationships between final and original raw scale scores for personality scales are contained with Table 5.23. Original and refined scale scores for integrity, conscientiousness, neuroticism, and dissimulation are all significant and strong.

Table 5.23

Personality Refined versus Raw Scale Scores

	Integrity Refined	Conscientiousness Refined	Neuroticism Refined	Dissimulation Refined
Integrity Raw Total	**.96***	-.44*	.33*	-.39*
Conscientiousness Raw Total	-.47*	**.94***	-.23*	.35*
Neuroticism Raw Total	.38*	-.31*	**.94***	-.33
Dissimulation Raw Total	-.45*	.50*	-.35*	**.93***

Note. Correlations marked * are significant at p < .007, N=317, (Listwise deletion of missing data).

Section 5.2.5: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability

This section presents results relevant to the present dissertation's first research objective. More specifically, it investigates whether previously observed relationships among integrity, conscientiousness, neuroticism, and ability are replicated. Wherever possible, it also explores variations among any such differences at the level of facet or ability subcomponent.

Section 5.2.5.1: Correlation Coefficients

Table 5.24 details the observed correlations between the scales piloted in this study. The observed correlation between the scales of conscientiousness and integrity was $r = -.50$, $p < .007$. This indicates that respondents who more strongly endorsed conscientious items were less likely to align themselves with behaviour associated with counterproductive workplace behaviour. The correlation between the scales of neuroticism and integrity was $r = .42$, $p < .007$. This indicates that respondents who more strongly endorsed neuroticism items were more likely to align themselves with behaviour associated with counterproductive workplace behaviour.

Table 5.24

Inter-scale Correlations amongst Crystallised Intelligence, Fluid Intelligence, Conscientiousness, Neuroticism, Integrity, and Dissimulation.

	Crystallised	Fluid	Conscientious	Neuroticism	Integrity	Dissimulation
Crystallised	1.00	**0.37****	**-0.13***	-0.10	0.08	**-0.21****
Fluid	**0.37****	1.00	-0.08	0.01	**0.16****	**-0.16****
Conscientious	**-0.13***	-0.08	**1.00**	**-0.38****	**-0.50****	**0.55****
Neuroticism	-0.10	0.01	**-0.38****	1.00	**0.42****	**-0.41****
Integrity	0.08	**0.16****	**-0.50****	**0.42****	1.00	**-0.46****
Dissimulation	**-0.21****	**-0.16****	**0.55****	**-0.41****	**-0.46****	1.00

Note. Correlations marked * are significant at p < .007, N=317, (Pairwise deletion of missing data).

While no significant relationship was found between integrity scores and outcomes on the crystallised intelligence scale, a small significant positive correlation was observed between fluid intelligence and integrity $r = .16$, $p < .007$. This suggests that there is a slight tendency for those with higher problem solving ability to answer integrity scale items in a direction associated with increased risk of counterproductive workplace behaviour.

No significant relationships were observed for ability indices and neuroticism scores. Nor were any significant relationships observed between ability and facet-level neuroticism traits. No significant relationship was observed between conscientiousness and fluid intelligence, but a small significant negative correlation was found with crystallised intelligence r = -.13, p < .007. This suggests that those who endorse items indicative of conscientiousness are slightly less likely to perform as well as others on ability items related to knowledge and education.

Section 5.2.5.2: Significance of Correlation Coefficient Differences

This section reports a calculation to determine the significance of differences among correlation coefficients between integrity and conscientiousness, and integrity and neuroticism. These computations were again based on Streiger's (1980) recommendations for both large sample sizes and Type I error control.

The calculation performed used significant correlation coefficients for integrity scores and the traits conscientiousness and neuroticism detailed in Table 2.24. The relationship between conscientiousness and integrity was significantly stronger than the relationship between neuroticism and integrity $t(314) = -11.80$, $p < .001$.

Section 5.2.5.3: Multiple Regressions

To more fully explore the unique variance in integrity explicable via conscientiousness and neuroticism *Forward Stepwise Regression* was undertaken (Gupta, 2000). Relevant output from regression model 1 is reported in Table 2.26 and suggests that the conscientiousness scale alone can explain 21% of integrity variance (β = -.460). There is 95% confidence that the squared multiple correlation is between – .497 and -.032.

Regression model 2 included the personality traits of conscientiousness and neuroticism. This model could account for 25.5% of variance in integrity scores. Both conscientiousness (β = -.386) and neuroticism (β = .231) were significant predictors. There is 95% confidence that the squared multiple correlation is between -.433 and -.254 for conscientiousness, and between .079 and .202 for neuroticism.

Regression model 3 accounted for 26.5% of variance in integrity scores. Both conscientiousness (β = -.362) and the neuroticism facet of mercurial (β = .260) were significant predictors. The neuroticism facet of angst was not found to significantly explain unique variance in integrity scores (β = -.143, p < .005). There is 95% confidence that the squared multiple correlations are between -.413 and -.231 for conscientiousness, and between .338 and .779 for mercurial.

Table 5.25

Beta values for Multiple Regression Coefficients of Personality Traits on Integrity

Trait	Integrity β
Conscientious	**-0.460***
Regression model 1	$F(1,315) = 74.760$
R	0.460
R^2	0.212
Adj. R^2	**0.210**
Conscientious	**-0.386***
Neuroticism	**0.231***
Regression model 2	$F(2,314) = 55.088$
R	0.510
R^2	0.260
Adj. R^2	**0.255**
Conscientiousness	**-0.362***
Neuroticism Angst	0.143
Neuroticism Mercurial	**0.260***
Regression model 3	$F(2,314) = 58.012$
R	0.519
R^2	0.270
Adj. R^2	**0.265**

Note. βs' marked * are significant at p< 0.007

Section 5.2.6: Adverse Impact Risk and Reduction for those identifying as Māori

This section presents results relevant to the present dissertation's second research objective. More specifically, it investigates whether there is a difference in personality or ability test scores between those identifying as Māori and non- Māori. It then explores whether any such differences could be reduced by combining ability tests with other assessments. As with the first study undertaken, the results detailed are not expected to resolve issues relating to adverse impact. The results are instead intended to add to the empirical foundations that may allow future assessment and selection practices to more radically reduce adverse impact.

Section 5.2.6.1: Mean Score Differences

Table 5.26 reports differences between those identifying as Māori and the broader participant group on the indices of crystallised (*d* = .40), fluid (*d* = .36), and general ability (*d* = .37). None of the differences observed were statistically significant. This may have been due to the very small N.

Table 5.26

Comparison of Māori and broader Participant Group Scale Means and Standard Deviations

Scale	Items	Māori Mean	Māori SD	N	Group Mean	Group SD	N	Cohen's *d*
Crystallised Ability	14	7.77	3.20	18	8.98	2.77	196	0.40
Fluid Ability	16	9.23	3.50	19	10.49	3.36	196	0.36
General Ability	30	17.62	6.74	22	19.99	5.91	196	0.37
Integrity	24	66.36	15.03	22	64.88	13.13	196	-0.10
Conscientiousness	27	95.86	14.79	21	97.86	15.26	196	0.13
Neuroticism	36	121.14	22.34	21	120.60	23.94	196	-0.02
Dissimulation	14	40.64	9.00	22	40.55	8.27	196	-0.01

Note. Differences would have been marked * for significant at $p < .05$, and ** for significant at $p < .001$, but no differences were significant.

Table 5.26 reports no significant or noteworthy difference on integrity scores between Māori and the broader participant group assessed (*d* = -.10). Nor any difference of note between Māori and the broader participant group for scores on the scales assessing neuroticism (*d* = -.02), or conscientiousness (*d* = -.13).

Section 5.2.7: Response Time

This section presents results relevant to the present dissertation's third research objective. More specifically, it investigates whether the length of time taken to answer ability assessments is related to respondent ability and item difficulty. It then explores whether differences in conscientiousness or neuroticism scores mean respondents take longer to answer ability assessments.

Section 5.2.7.1: Response Time, Ability, and Item Difficulty

Table 5.27 reports relationships between the average length of time taken to complete all items within an ability scale and scores on ability scales. Small negative relationships were observed between crystallised intelligence response latency and crystallised intelligence scores (r = -.21, p <.001), and IRT derived respondent ability level (r = -.24, p <.001). Fluid intelligence indices of ability were also significantly correlated with mean fluid intelligence latency. However, Table 5.27 shows that fluid intelligence latency had a small *positive* relationship with total fluid intelligence scores (r = .18, p< .001), and IRT-derived respondent ability level (r = .17, p < .001).

Table 5.27

Relationships between Respondent Ability and Response Latency

	Crystallised Latency	Fluid Latency
Crystallised Intelligence Weighted Total	**-.21***	.01
Crystallised Intelligence Ability Level (IRT)	**-.24***	-.03
Fluid Intelligence Weighted Total	-.01	**.18***
Fluid Intelligence Ability Level (IRT)	-.01	**.17***

Note. Correlations marked * are significant at p < .007, N=317, (Listwise deletion of missing data).

To test for potential obfuscation, based on non-linear relationships, ability levels and response latencies were further explored via one-way ANOVAs. Eight IRT-derived respondent ability level groups were created. These groups ranged in ability level from > -1.5 through to > 1.5 and categorisation occurred according to .5 increments. The effect of crystallised intelligence ability was not significant for crystallised intelligence latency, F(7, 309) = .704, p = .669. Nor was the effect of fluid intelligence ability significant for fluid intelligence latency, F(7, 309) = .587, p = .767.

The relationships amongst crystallised intelligence item latency indices and crystallised intelligence item difficulty indices are reported in Table 5.28. Moderate negative relationships were observed between the mean response time for items for those correctly answering questions and the proportion of respondents correctly answering a question (r = -.68, p <.005), and the log of correct responses divided by incorrect/error responses (*r* = -.67, *p* <.006). The standard deviation of response latencies for those answering items correctly was also moderately correlated with the item difficulty indices of proportion correct (*r* = -.71, *p* <.003) and the log of correct responses divided by incorrect responses (*r* = -.70, *p* <.003).

Moderate negative relationships were also observed between an item's mean response time for those answering correctly minus those incorrectly responding and the proportion of respondents correctly answering a question correct (*r* = -.75, *p* <.001), and the log of correct responses divided by incorrect/error responses (*r* = -.74, *p* <.001). The standard deviation of response latencies for those answering items correctly divided by those answering incorrectly was also moderately correlated with the item difficulty indices of proportion correct (*r* = -.70, *p* <.003) and the log of correct responses divided by incorrect responses (*r* = -.71, *p* <.003).

Table 5.28

Relationships among Crystallised Intelligence Item Difficulty Indices and Response Latencies

Crystallised Latency Measures	Proportion Correct	Log C/E
Group Mean	-.56	-.55
Group Standard Deviation	-.53	-.53
Correct Response Mean	**-.68***	**-.67***
Correct Response Standard Deviation	**-.71***	**-.70***
Incorrect Response Mean	-.08	-.07
Incorrect Response Standard Deviation	.02	.04
Correct-Error Mean	**-.75***	**-.74***
Correct-Error Standard Deviation	**-.70***	**-.71***

Note. Correlations marked * and emboldened are significant at p < .007, N=317, (Listwise deletion of missing data). Proportion correct and the Log of correct divided by incorrect responses serve as the discrimination indices.

A number of moderate to strong significant relationships were found between fluid intelligence latency indices and fluid intelligence item difficulty indices. These are reported in Table 5.29. Strong negative relationships were observed between an item's mean response time and the proportion of respondents correctly answering it (*r* = -.85, *p* <.001) and the log of correct responses divided by incorrect/error responses (*r* = -.85, *p* < .001). The standard deviation of response latencies for the participant group was also moderately correlated with the item difficulty indices of proportion correct (r = -.64, p < .006) and the log of correct responses divided by incorrect responses (r = -.65, p < .005).

Strong negative relationships were also observed between the mean item response time for correct answers and the proportion of respondents providing them (r = -.86, p < .001) and the log of correct responses divided by incorrect/error responses (r = -.86, p <.001). The standard deviation of response latencies for the participant group was also moderately correlated with the item difficulty indices of proportion correct (r = -.66, p <.003) and the log of correct responses divided by incorrect responses (r = -.67, p <.003).

Strong negative relationships were also observed between the mean item response time for those correctly answering questions minus the those incorrectly responding and the proportion of respondents correctly answering a question correct (r = -.81, p <.001) and the log of correct responses divided by incorrect/error responses (r = -.82, p <.001).

Table 5.29

Relationships among Fluid Intelligence Item Difficulty Indices and Response Latencies

Fluid Latency Measures	Proportion Correct	Log C/E
Group Mean	**-.85***	**-.85***
Group Standard Deviation	**-.64***	**-.65***
Correct Response Mean	**-.86***	**-.86***
Correct Response Standard Deviation	**-.66***	**-.67***
Incorrect Response Mean	**-.81***	**-.82***
Incorrect Response Standard Deviation	-.44	-.45
Correct-Error Mean	-.27	-.26
Correct-Error Standard Deviation	-.04	-.03

Note. Correlations marked * are significant at p < .007, N=317, (Listwise deletion of missing data). Proportion correct and the Log of correct divided by incorrect responses serve as the discrimination indices.

Section 5.2.7.2: Response Time and Personality Traits

Correlation coefficients used to examine relationships amongst crystallised and fluid intelligence latencies and conscientiousness and neuroticism scores are detailed in Table 5.30. Neither of the relationships between crystallised intelligence and conscientiousness or neuroticism was found to be significant. No significant relationship was found between neuroticism and fluid intelligence response latency either. A relationship between conscientiousness and fluid intelligence was observed (*r* = -.16, *p* < .005). Although significant, the strength of this relationship does not reach Ferguson's (2009) minimum recommended effect size of .2.

Relationships between conscientiousness and neuroticism scores and latency were also examined for each fluid and crystallised intelligence item in case mean level relationships were obfuscated. A significant relationship was observed for fluid intelligence item 21 latency and neuroticism (*r* = -.17, *p* < . 002). No other significant relationships were found between crystallised or fluid intelligence item latencies and neuroticism or conscientiousness scores.

Table 5.30

Relationships among Conscientiousness, Neuroticism, and Mean Response Latencies on Ability Scales

	Gc Latency	Gf Latency
Conscientious	.08	**.16***
Neuroticism	-.03	-.09

Note. Correlations marked * and emboldened are significant at p < .007, N=317, (Listwise deletion of missing data).

Relationships were then examined between crystallised and fluid intelligence item mean latency and the two distinct neuroticism factors identified during factor analysis. These relationships are detailed in Table 5.31. A small significant relationship was identified between neuroticism mercurial factor and fluid intelligence item mean latency (*r* = -.17, *p* < .002). Once again, item-level latency and personality scores were also examined. Fluid intelligence item 10 had a significant relationship with neuroticism mercurial factor (*r* = -.18, *p* < .001). Fluid intelligence item 21 also had a significant relationship with neuroticism mercurial factor (*r* = -.19, *p* < .001). Fluid intelligence item 24 also had a significant relationship with the mercurial factor (*r* = -.15, *p* < .006).

Table 5.31

Relationships among Neuroticism Factors and Response Latencies on Ability Scales

	Crystallised Latency	Fluid Latency	Crystallised Score	Fluid Score
Angst Factor	-.02	-.02	-.01	.01
Mercurial Factor	-.08	**-.17***	-.03	-.02

Note. Correlations marked * are significant at p < .007, N=317, (Listwise deletion of missing data).

Section 5.3 discusses potential interpretations, implications, and limitations of the preceding findings. The degree to which these results have answered questions posed from the preceding study will also be examined.

Section 5.3: Discussion (Study 2)

This section discusses the preceding results. Section 5.3.1 examines findings in reference to the first research objective. This includes a brief discussion of consistency with previous research. Further discussion on the consistency of Study 2's findings with those of Study 1 is undertaken in the following integration chapter (Chapter 6). Section 5.3.2 repeats this process for the second research objective. Findings relevant to the third research objective are comprehensively discussed in Section 5.3.3. Speculation as to why findings may have occurred and a discussion of the consequent implications is also undertaken. Section 5.3.4 then discusses the outcomes of assessment development. Specific limitations and directions for future research are then examined. More general limitations and future directions are detailed in Chapter 6.

Section 5.3.1: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability.

In line with the first research objective, this section investigates the convergence of results with previously observed relationships among integrity, conscientiousness, neuroticism, and ability. Where possible, it then explores variations among any such differences at the level of facet or ability subcomponent.

There was a strong relationship between increases in conscientiousness and increases in integrity. This is consistent with a large body of international research (e.g., Ones, 1993, 2003; Ones et al., 1993a, 1995; Sackett & Wanek, 1996; Wanek, 1995, 1999). Due to the conscientiousness scale only comprising a single trait-level it was not possible to investigate any differences in relationships at the facet level.

Neuroticism was also found to have a strong relationship with integrity. Respondents displaying greater neuroticism consistent behaviour were more likely to report tendencies associated with counterproductive workplace behaviour. The strength of this relationship is consistent with previous investigations (e.g., Ones, 1993, 2003; Ones et al., 1993a, 1995). The item content of the neuroticism scale employed in this investigation may have important implication for any explanations of this research. One of the common facets of neuroticism is suspicion or cynicism. The present investigation's neuroticism scale lacked questions capturing this suspicion/cynicism facet (see Appendix A). This suggests that the consistent cognitions often associated with neuroticism (e.g., people are fundamentally not to be trusted) may be less important to explaining the integrity and neuroticism relationship than its more transitory emotional reactions (e.g., losing one's temper). If further supported, this possibility suggests that the display of counterproductive behaviour may be more situational then presently captured by standard integrity items. Standard overt integrity items use historical incidences of counterproductive behaviour and enduring anti-social ideation to screen out high risk applicants. They do not account for counterproductive behaviour influenced by more transitory emotional reactions. Applicants without a propensity towards CWBs, but with tendencies towards volatility, may answer integrity items accurately, but still pose a high risk of counterproductive behaviour in some situations. Example situations could include a workplace culture that is hard-driving, pressured, and highly task-focused. An alternative example could be if such employees experience some personal crisis that lowers their threshold for coping with stress and pressure.

The observation that integrity scores have slightly stronger relationships with conscientiousness than neuroticism is consistent with previous findings (Ones, 1993, 2003; Ones et al., 1993). The moderate to strong relationships amongst integrity, conscientiousness, and neuroticism also provide further tentative support for Digman's (1997) theory that super-ordinate Alpha factor maturation influences the development of impulse control and conscience. This implication was discussed extensively within the preceding chapter.

The present study found a small relationship between integrity and fluid ability. This relationship suggested that small increases in problem solving capacity were associated with small increases in the risk of counterproductive behaviour. The idea that those with more innate intellectual potential are more likely to engage in counterproductive behaviour or harbour enabling attitudes towards counterproductive behaviour is inconsistent with previous findings (e.g., Cullen & Sackett, 2004; Ones & Viswesvaran, 1998a; Sackett & Wanek, 1996). It also runs counter to reviews that strongly link lower intelligence and criminal offending (Herrnstein & Murray, 1994). The content of some integrity items may provide a more likely explanation for this contradictory finding. Many of the integrity items upon which this correlation was based were written to maximise Greenberg's (1997) geometry of theft rationalisations. This involved providing a hypothetical context in which rationalisation and justification were possible. For example, "Employees often steal because they are not treated fairly by their boss." It may be that people with relatively higher problem solving capability can discern the likelihood that such rationales are applied by those at risk of counterproductive behaviour. This recognition could result in more capable problem solvers affirming such items without implying the standard heightened risk of CWBs. The small strength of this relationship provides further support for an explanation that involves a tendency to answer some integrity items in a particular way, but doesn't increase the general tendency to affirm all integrity items.

No significant correlations were found between neuroticism and any of the ability indices. This is inconsistent with previous findings that ability and neuroticism are negatively correlated (Ackerman & Heggestad, 1997; Austin et al., 2002; Hembree, 1988; Kyllonen, 1997; Seipp, 1991). The most likely explanation for this divergence again focuses on disparities in item content regarding the suspicion/cynicism facet. This explanation gains considerable support from comparisons with Study 1 findings and will be a subject of more comprehensive discussion in the following integration chapter.

Conscientiousness was found to have a small negative correlation with crystallised ability, but no significant correlations with fluid ability. This finding is consistent with previous research examining the relationship between conscientiousness and ability indices within a New Zealand participant group (Wood & Englert, 2009). However, it is inconsistent with previous U.K.-based findings that suggested conscientiousness is negatively correlated with fluid, but not crystallised ability (Moutafi, et al., 2003, 2004). The explanation for this difference concerned participant characteristics and was comprehensively discussed in Chapter 4.

Section 5.3.2: Adverse Impact Risk and Reduction for those identifying as Māori

No differences between Māori and non-Māori for ability or personality scores were statistically significant. On this basis no adverse impact calculations or potential reductions were explored.

Section 5.3.3: The relationship between response time, respondent ability, and item difficulty.

In line with the third research objective, this section discusses whether the length of time taken to answer ability assessments is related to respondent ability or item difficulty. It then explores whether differences in conscientiousness or neuroticism scores meant respondents took longer to answer ability assessments.

Respondent ability was found to be related to the length of time it takes to complete ability scales. The average response time for completion of the crystallised ability scale significantly decreased as respondent ability increased. However, there was also a significant positive correlation between response time and respondent ability for the fluid scale. In other words, those scoring higher on fluid ability had significantly *longer* total average response times than those scoring less well. That more knowledgeable respondents answer crystallised items more quickly appears consistent with the historical view that speed of problem solving is an important aspect of intelligence (Thorndike, 1913; McFarland, 1930) and ability (Glickman et al., 2005; Scrams & Schnipke, 1997). The finding that those with greater problem solving capability took longer to answer fluid items appears inconsistent with this speed hypothesis. It is not the first time that a positive relationship between ability scores and the length of time taken to complete items has been observed (Entink, Fox, & van der Linden, 2009; Scrams & Schnidke, 1997). Scrams and Schnipke (1997) suggested that the incongruence of their results with the belief that higher ability examinees process information more quickly may have been due to a confound resulting from the timed nature of the assessment, and the fact that not all respondents used the time available or completed all the items. This explanation is further supported by findings that respondents often switch from solution-based behaviour to rapid guessing for items towards the end of an assessment with an overall time limit (van der Linden, 2009; Wise & DeMars, 2006). In such circumstance lower scores, but quicker overall response times, may occur due to relatively rapid guessing towards the end of an assessment. One problem with looking to such a time-management confound as an explanation for the current investigation's findings is that time-management issues were reduced through providing item-by-item time limits rather than an overall time limit. On this basis there was less reason for respondents to switch from solution-based behaviour in earlier items to rapid guessing behaviour for later items.

A more likely explanation is that the positive and negative relationships between RT and fluid and crystallised ability are a direct result of an interaction between respondent ability and the cognitive requirements associated with correctly responding to these items. As previously discussed, crystallised items are primarily solved on the basis of retrieving and using prior knowledge. On this basis providing the correct answer may require considerably less time for those who already have the background knowledge necessary to do so (Yang, 2007). On the other hand, fluid items are specifically designed to reduce and minimise the impact of prior knowledge on correctly responding. Fluid items are designed to require complex processing rather than simple retrieval in order to find the right answer. This explanation is also consistent with the differences in mean response times for fluid and crystallised items. Table 5.3 shows that the average response latencies for the 30 crystallised items originally piloted ranged from 6.54 to 17 seconds. Table 5.4 shows that the average time taken for the 30 fluid items originally piloted ranges from 14.02 to 32.77 seconds. These differences between average response times for fluid and crystallised items are indicative of the different processing requirements associated with providing a response. This contention is further supported by the total item information functions detailed in Figure 5.7 and Figure 5.8. These curves suggest that despite taking considerably longer to answer on average, the fluid items were actually slightly easier than the crystallised items. This appears to reduce the possibility that the difference in response times between fluid and crystallised items is attributable to their difficulty rather than processing requirements.

This difference in cognitive requirement explanation is also broadly consistent with more general findings within response time research. Frearson and Eysenck (1986) introduced the *odd-*

man-out paradigm within simple decision-making tasks (Hick tasks) in order to increase the information-processing demand of correct responses. Their subsequent results suggested that response time increases in line with information-processing demands. This is further supported by the observation that ability tasks take differing amounts of time to complete based upon their processing demands (Jensen, 1998; van der Linden, 2007).

The difficulty of ability items was also found to have a relationship with how long it takes respondents to answer those items. Easier questions across both fluid and crystallised items took respondents less time to answer than more difficult questions. This is consistent with Neubauer (1990) and Verbic's (2010) findings. However, it is interesting to note that the strength of this relationship varied across fluid and crystallised items. Moderate to strong relationships were observed between the difficulty of crystallised items and the length of time respondents took to provide correct responses, but not how quickly respondents tended to answer these items in general. In this respect the average amount of time spent trying to answer crystallised items had no significant relationship with their difficulty or ease. On the other hand, for fluid items there was a strong significant relationship between how long respondents tended to take attempting to answer items and how difficult they were. Providing correct or incorrect answers appeared to have little impact on the length of time taken to respond.

Such findings appear problematic for the generally held assumption that the probability of correctly responding is a monotonically increasing function of RT with both a lower and upper asymptote (Entink et al., 2009; Schnipke & Scrams, 1997; Scrams & Schnipke, 1997; van Breukelen, 2005). The probability of a correct response has traditionally been seen as increasing and then stabilising as a function of the amount of time spent answering an item (Wang & Hanson, 2005; Wang & Zhang, 2006; Yang, 2007). This is not consistent with the finding that people who answer crystallised type items correctly do so faster than those who respond incorrectly.

The origins of response time modelling provide an explanation for the lack of correct-response probability increasing as a monotonic function of RT. The modelling behind this expectation is based upon cognitive and experimental psychological research involving speeded elementary cognitive tasks (van der Linden, 2007; Wenger, 2005). In other words, it is based upon research involving items that are solvable given sufficient time. Within this context incorrect responses are due to time pressure (van Breukelen, 2005). Items within power assessment such as that used in the present investigation instead cover a range of greater complexity and are not all solvable by all examinees – regardless of time. The remainder of this section will look at the impact of conscientiousness and neuroticism on the length of time taken to answer ability scales.

Those scoring higher on the conscientiousness scale tended to take slightly longer to answer fluid items than others. However, no relationship was found between conscientiousness and the length of time taken to answer crystallised items. The fluid ability finding is consistent with findings of a positive relationship between conscientiousness and time spent on a task (Biderman, Nguyen, & Sebren, 2008). Biderman et al.'s methodology operationalised "time-on-task" as number of hours spent studying for a test and other preparatory related behaviours. However, the same self-regulation theories posited to understand the observed relationship may shed light on this investigation's findings. According to Bandura's (1991) social cognitive theory of self-regulation, people are motivated by a self-regulatory mechanism that includes judging one's behaviour regarding both personal standards and situational constraints. In the ability test performance context of this investigation, the self-regulatory process may have a stronger influence on conscientious respondents' commitment to trying to answer an item correctly and subsequently spending additional item on fluid items in order to achieve this goal.

The finding that more conscientious respondents take longer to answer fluid items may provide a partial explanation for previous research findings that conscientiousness appears negatively related to fluid ability, particularly on high-difficulty assessments designed for use on professionals, managers, and/or graduates (e.g., Moutafi et al., 2004). Bridgeman (1980) theorised that respondents adopting a slow and careful approach to answering items in a timed testing condition may end up penalised for not finishing the assessment. It is also possible that conscientious respondents answer as many questions as their more expedient counterparts, but that they are prone to adopting less adaptive pacing strategies and therefore score less well (e.g., spending too long on early items and then guessing a higher proportion of items towards the end of the assessment). During an assessment of moderate difficulty the additional time spent on answering earlier fluid items may result in a similar overall score, but with relatively more correct answers provided at the beginning of the test. However, during an assessment of higher difficulty devoting additional time to items at the beginning of the assessment may not actually translate into more correct answers. This is because more items within a higher difficulty assessment are likely to be unanswerable by a higher proportion of respondents – regardless of time spent seeking a solution. Although these scenarios and the preceding explanations may account for differences in outcomes within overall timed testing environments, they cannot account for the present investigation's finding that crystallised ability and conscientiousness were negatively related. The present investigation provided time-limitation information for each item rather than simply for the test overall. This is likely to have facilitated more item-specific pacing strategies, and therefore reduced the likelihood of maladaptive responding as time ran out for latter items.

Unlike conscientiousness, neuroticism was not found to be related to the length of time respondents took to complete ability scales. As previously mentioned, the existence of a negative relationship between ability response times and neuroticism would have been consistent with the neural efficiency hypothesis, which suggests respondents experiencing relatively high levels of anxiety engage in substantially more task irrelevant processing (i.e., worry) (Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Zeidner & Matthews, 2000). A simple explanation for this failure to observe results consistent with this hypothesis concerns the fact that the low stakes nature of this project reduced the likelihood that respondents would be as concerned about their results. The item by item timing may also have gone some way to reducing the sense of pressure experienced by respondents.

However, when the neuroticism scale was split into its facet-level traits it was observed that those endorsing the "mercurial" factor tended to take less time answering fluid ability items than other respondents. That the relationship with mercurial was observed for fluid but not crystallised items is consistent with Moutafi et al.'s (2005) finding that neuroticism can explain more variance in fluid than crystallised ability indices. Again, this is explained as a function of processing requirements, wherein fluid ability items require greater processing resources than crystallised items and are therefore more prone to neurotic interference. Yet the finding that only the mercurial facet had a significant relationship is interesting. Based upon the processing interference hypothesis of Moutafi et al. it could reasonably be expected that the tendency to worry and more acutely experience anxiety would correlate with longer response times. The finding that the angst facet had no relationship with how long it took respondents to answer fluid items runs counter to this expectation. The results instead suggested that performance on fluid ability may be confounded by an increased propensity to give up in frustration rather than continue with solution seeking behaviour. The following section discusses the success of scale piloting and the implications of research objectives for assessment design and use.

Section 5.3.4: Composite Assessment Development Outcomes and Implications

Scales measuring integrity, conscientiousness, neuroticism, and dissimulation were developed, piloted, and assessed within the statistical framework of classical test theory. The final items comprising each of these scales demonstrated acceptable item discrimination, unique factor loadings of appropriate strength, and avoided excessive thematic redundancy. The factors identified via EFA were in line with the overarching scales intended. The final personality scales demonstrated no noteworthy differences on the basis of gender or ethnicity.

The scales measuring crystallised and fluid ability also arose as unique factors within EFA modelling. The items comprising these scales were selected in line with classical test theory analyses. The items retained demonstrated sufficient difficulty, discrimination, and content range. The accompanying distractors also demonstrated desired properties. In particular, they were more likely to be chosen by those in the lower one-quarter of overall scorers on each ability scale. Retained ability items were then analysed via a 3-parameter IRT model. This analysis suggested these items had acceptable difficulty, discrimination, and likelihood of guessing parameters. It also suggested that the overall difficulty of each ability scale was approximately consistent with assessing average ability respondents. This was consistent with the level of difficulty originally intended.

There scales of conscientiousness, neuroticism, and integrity were found to have moderate relationships. The scales developed are for constructs that are well-substantiated predictors of future job performance (e.g., Schmidt & Hunter, 1998). However, the size of their interrelationships indicates their combined predictive accuracy is likely to be considerably less than the sum of their parts. The finding that conscientiousness had a negative relationship with crystallised ability does not raise this same concern due to the small strength of this relationship. That said, this relationship does suggest that the accurate prediction of job performance may be enhanced by exploring the potentially compensatory nature of the crystallised ability and conscientiousness relationship (Wood & Englert, 2009). This is further supported by the finding that the negative relationship between conscientiousness and crystallised ability cannot be explained as a confound related to more conscientious respondents taking longer to answer crystallised items.

Response time outcomes also had an important implication for current and future test design. Previous research suggested that the length of time taken by a respondent to answer ability items could potentially be used as one estimate of their ability (Glickman, Gray, & Morales, 2005; Scrams & Schnipke, 1997). In other words, those correctly answering ability items relatively quickly could be given higher scores. The small size of relationships between ability scores and the length of time taken to complete ability scales suggests that awarding respondents extra points on the basis of the speed in which they provide correct answers is not appropriate. That the direction of the relationship between response time and scale outcomes differed across fluid and crystallised ability makes the awarding of extra points even more inappropriate. Those scoring higher on fluid ability tended to take longer to complete this scale. If extra points were awarded on the basis of this relationship then those who took relatively *longer* to correctly answer questions would gain higher scores. The next section examines some of the specific limitations of this study and its results. More general limitations are discussed in Chapter 6.

Section 5.3.5: Specific Limitations

A limitation of this investigation's design was its failure to comprehensively assess the validity and reliability of the scales piloted. This study was able to provide tentative reliability support through an examination of the internal consistency of items and scales. However, it did not provide any assessment of test-retest consistency. The use of Cronbach's α as an estimate of reliability was also potentially problematic. A number of recent studies and reviews have suggested Cronbach's α tends to underestimate reliability (Bentler, 2009; Green & Yang, 2009; Sijtsma, 2009a; Liu, Wu, & Zumbo, 2009).

This study was able to provide tentative support for construct validity through EFA and discriminant comparison amongst scales. However, it did not examine the convergent validity of scale scores with other recognised and validated scales of the same constructs. Nor did it examine the predictive or incremental validity of the scales if used to predict job performance individually or in combination.

The accuracy of ability item parameters is another potential limitation within this study. The validity of ability estimates is dependent upon the amount of effort expended by respondents when completing the assessment. If adequate effort is not made then ability estimates are likely to be underestimates and (Wise & DeMars, 2006). Furthermore, motivated examinees significantly outperform their less motivated peers (Wise & DeMars, 2005). The assessment piloting within this dissertation occurred under low-stakes circumstances for respondents. Despite the removal of cases displaying suspiciously rapid responding prior to analysis, the concern that respondents may not have displayed maximum effort on ability items remains an important one regarding the accuracy of item parameters. Yang (2007) found the disproportionately common guessing behaviour displayed in low-stakes situations to result in item parameter distortions. More specifically, Yang found guessing to result in the overestimation of discrimination parameters for easy items and underestimation of discrimination parameters for more difficult items. Rapid guessing has also been found to lead to an overestimation of item difficulty and an underestimation of the pseudo-guessing parameter (Meyer, 2008).

Another potential limitation of this investigation concerns the analysis of response time. The analysis undertaken was relatively simple. Other researchers have suggested it is beneficial to take into account additional information when analysing response time information. Such additional information includes the attributes of shift, scale, or shape (Rouder, 2005a). Another limitation regarding response time analyses was the failure to control for the variable of age or gender. Jensen (1998) found test-taking speed and RT to decline as a function of age, which is likely to increase the effect of correlations for speeded tests amongst elderly people. Recent research by Verbic (2010) looked at gender differences for the pace adopted when answering items in a low-stakes setting. Verbic found a statistically significant difference between male and female total response times ($t = 6.6$, $p < .001$), but an examination of RT distributions suggested this was due to females having a longer warm-up effect. In other words, they took longer to answer items at the beginning of the assessment.

Only collecting per item response time also limited the findings of this study. Had a second participant group encountered the same items within an overall assessment timing condition it would have been possible examine the impact per item timing had on pacing strategies.

Another crucial limitation within this study was the relatively small participant group. Factor analysis and item response theory both require extremely large participant groups to meet underlying assumptions and have confidence in results. The recommended participant group size for the exploratory factor analysis methodologies used is sometimes as large as ten respondents for every variable (DeCoster, 1998). The largest EFA undertaken involved 112 personality items (conscientious, neuroticism, integrity, and distortion). Using the ten to one ratio a participant group of 1120 respondents would have been optimal. The actual participant group was 317 respondents. While this falls considerably short of the putatively optimal 1120 number, it still exceeded the 100 subject minimum recommended by Barrett and Kline (1981) and more conservative 200 minimum recommended by Guilford (1956). The relatively small participant group may also limit the accuracy of item characteristic curves and total information functions. Having a sufficiently large participant group is very important to correctly identify the of accuracy item parameter estimates in IRT (de la

Torre & Hong, 2010). Most IRT practitioners agree that many hundreds of respondents are required for confidence in the outcomes of analysis. The commonly cited number is between 500 and a 1000 respondents (Thissen, 1991), although with smaller participant groups a reasonable degree of precision can be expected (Stark & Chernyshenko, 2006).

The next section will examine some potential future directions for subsequent research-based upon these limitations and other considerations. More general directions for future research are detailed in Chapter 6.

Section 5.3.6: Future directions

This study's first goal was to provide tentative empirical support for the psychometric properties of ability, integrity, conscientiousness, and neuroticism scales. Considerable work is still required to increase confidence in the psychometric properties of these scales. Test-retest analysis must be undertaken for reliability. Alternatives to Cronbach's chould be explored (Revelle & Zinbarg, 2009); although the tools required to perform calculations for these alternatives are so esoteric as to render them unlikely alternatives in their current form (Sijtsma, 2009b). There would also be considerable value in seeking future support for the construct validity of scales through an analysis of similarities between scale scores on these and extant scales of similar constructs. A number of other post-doctoral activities are also required before the composite designed could be utilised within a selection context. For example, decisions around the usefulness of different weighting approaches for the component parts must be made. Options include the use of empirical regression weights, weights based on the judgement of subject matter experts, weights based on archival information (e.g., job analysis), and unit weights. The benefits and limitations of these respective methods are discussed in Bobko, Roth, and Buster (2007). A comprehensive predictive validity study involving job performance data must also be undertaken.

Another profitable area for future investigation relates to this investigation's use of non-applicants in a low-stakes setting. In other to avoid potential item parameter distortions future research should retest the current parameters within a high-stakes setting. It should also look to involve a participant group large enough to counter the previously discussed size limitations within the current investigation. There would also be benefit in more sophisticated analyses of response time information and the use of control versus experimental groups to test RT related hypotheses.

Another direction for future research will be the use of Tatsuoka's (2009) *rule-space model* of test design. The commonly used classical test theory and IRT frameworks focus exclusively on response accuracy (Scrams & Schnipke, 1997). Performance is measured for the number (CTT) or characteristics (IRT) of the items answered incorrectly or correctly. Tatsuoka's (2009) rule-space model is instead a statistical method for aligning test taker item responses with a set of attributes associated with different cognitive skills and skill-levels. The basic idea behind the rule-space's use within psychometrics is that tests are based on cognitive problem-solving tasks, which suggests cognitive psychology has the potential to advance psychometric theory. According to the rule-space model a respondent's ability to answer items correctly depends on whether or not they possess a specific set of cognitive skills or competencies called attributes. More specifically, an attribute of a task is a description of the knowledge, processes, or skills a test-taker would need to have in order to successfully complete a task. Knowing the specific attributes required to correctly answer items enables one to provide very specific information on their problem-solving strengths and areas for development. Such information also enables a prediction of exactly which type of items a particular testtaker should be able to answer correctly. This in turn facilitates very specific inferences about more general problem solving ability (Gierl, Leighton, & Hunka, 2005; Tatsuoka, 2009). The relevance of understanding the cognitive processes that lie behind success item responses is also now recognised as a crucial component of understanding the interaction of response times and ability estimates (Daniel & Embretson, 2010). The future use of this rule-space model may also facilitate testing the theory that RT and ability score relationships vary across fluid and crystallised indices as a function of the cognitive requirements associated with correct responses.

Section 5.3.7: Summary

This chapter investigated the three research objectives outlined in Chapter 3. The first of these suggested that previously observed relationships among integrity, conscientiousness, neuroticism, and ability were largely replicated within New Zealand. Integrity scores were again found to be related to conscientiousness and neuroticism (e.g., Ones, 1993, 2003; Ones et al., 1993a; Sackett & Wanek, 1996). Furthermore, the previously observed negative relationship between crystallised ability and conscientiousness was repeated (Wood, 2004; Wood & Englert, 2009). The lack of significant relationships between neuroticism and ability indices was the only substantial failure to replicate previous findings (e.g., Ackerman & Heggestad, 1997; Austin et al., 2002). The low-stakes nature of the research undertaken was offered as an explanation for this.

In investigating the second research objective no significant differences were found between those identifying as Māori and non-Māori on ability indices, conscientiousness, neuroticism, or integrity scales. However, the very small number of Māori participants (N = 22) reduces the statistical robustness of this finding.

The third research objective investigated whether the length of time taken to answer ability assessments was related to respondent ability or item difficulty. It was found that the length of time taken to answer ability scales was weakly and inconsistently related to scores on those assessments. It was also found that the length of time taken to answer questions increased as the difficulty of those questions increased. The third research objective also explored whether differences in conscientiousness or neuroticism meant respondents took longer to answer ability assessments. Differences in conscientiousness were weakly related to how long it took respondents to answer fluid items, but not crystallised items. Neuroticism had no relationship with the length of time taken to complete ability scales. However, the neuroticism facet of mercurial did have a negative relationship with the length of time taken to complete the fluid ability scale. In other words, those who scored as more mercurial tended to complete the fluid ability scale *faster*.

A composite assessment of integrity, conscientiousness, neuroticism, and fluid and crystallised ability scales was also developed as a function of this chapter. These scales and their items met the required parameters for psychometric legitimacy. More specifically, items associated with each of these scales demonstrated appropriate discrimination, difficulty, and distractor functioning. The scale-level analyses also suggested that the collections of items did indeed measure underlying factors consistent with expectations and did so with the required level of consistency. Furthermore, differences between males and females, and Māori and the broader participant group were minimal. The only major difference between the findings reported in this chapter and those expected on the basis of Chapter 2's literature review concerned facet-level traits. The literature review suggested considerable support for the existence of Big Five personality traits comprised of distinct facet-level traits. This investigation found no indication of distinct facet-level traits associated with conscientiousness, and only two facet-level traits associated with neuroticism. This is despite the author's attempt to comprehensively cover the relevant construct domains during item writing. The next chapter provides an integration of results from Study 1 and Study 2, it also details more general limitations and directions for future research.

Chapter 6: Integration and Conclusions

The purpose of this chapter is to briefly examine how well this dissertation achieved its research objectives. The discussion sections of Chapter 4 and Chapter 5 have already comprehensively addressed most findings and the majority of this discourse will not be repeated. The current chapter will instead focus most strongly on inconsistencies across Study 1 and Study 2. It will also examine some general limitations, future directions, and themes that have emerged across both studies.

Section 6.1: Relationships among Integrity, Conscientiousness, Neuroticism, and Ability.

This dissertation's first research objective was prompted by the need to better understand relationships among integrity, conscientiousness, neuroticism, and ability. Examining these relationships has clarified the construct domains of these predictors. It is hoped this will assist practitioners use these scales to make more accurate predictions of future job performance. It has also facilitated calculations regarding reductions in adverse impact by combining predictors. This is a focus of discussion in Section 6.2.

Study 1 and Study 2 both found integrity and conscientiousness scores to be related. More specifically, both studies suggested that self-reports of counterproductive workplace behaviour, more general dishonesty, and potentially enabling attitudes towards CWBs were less likely amongst those endorsing conscientious items concerning the tendency to organised, self-disciplined, goal-directed, and considered behaviours.

The biggest difference between Study 1 and Study 2 for this finding concerned the strength of association. Study 1 reported a weak relationship where conscientiousness could account for 6% of the variance in integrity test scores. Study 2 reported a moderate to strong relationship where conscientiousness could account for 21% of the variance in integrity scores.

If the first study had reported the stronger relationship it would be possible to attribute this difference to domain overlap between the items of the SSI and the 15FQ+'s conscientiousness scales. This would have been understandable as these are assessments measuring arguably overlapping constructs and were designed at different times by different people. However, the second study reported the strongest relationship and the integrity and conscientiousness items were selected on the basis of their unique factor loadings within shared exploratory factor analyses. For this reason a difference in construct domain focus between conscientiousness measures may be a more likely explanation for the inconsistency in findings.

Although the author did not have access to any mapping information regarding 15FQ+ items and the domains they load upon, some basic information is known. The conscientiousness facets in Study 1 each comprise 12 items putatively loading on characteristics associated with perseverance/responsibility, social reserve/self-monitoring, and fastidiousness/compulsiveness (Psychometrics Ltd., 2002). The link between these characteristics and the Big Five domain of conscientious are apparent, but the 15FQ+'s use of these facets to represent Big Five conscientiousness is a case of *ex-post facto* modelling. The facet level model already existed and was then subjected to additional rotation. The link is therefore more tenuous than the domain coverage and focus of Study 2's conscientiousness items, which were written from the outset to sample this domain. This explanation is also supported by the greater similarity between the strength of relationships reported in Study 2 and international research (e.g., Ones, 1993, 2003).

Study 1 and Study 2 both found that integrity was related to neuroticism. This result was consistent with international findings that integrity is negatively associated with neuroticism (Ones, 1993, 2003). Study 1 reported that among neuroticism facets, those pertaining to suspicion/distrust/cynicism and sensitivity to criticism/difficulty coping with stress explained all unique variance in integrity accounted for by neuroticism. On this basis it was argued that more cynical individuals with less capacity to cope with pressure and criticism were more prone to the perceptions of victimisation or more general strain associated with counterproductive behaviour (Chen, & Spector, 1992; Fox & Spector, 1999; Fox, Spector, & Miles, 2001; Yang & Diefendorff, 2009). That Study 2's neuroticism scale lacked facets or items capturing suspicion/cynicism may indicate that the consistent cognitions associated with neuroticism (e.g., people are fundamentally not to be trusted) are less important to explaining the integrity and neuroticism relationship than that more transitory emotional reactions (e.g., losing one's temper).

Study 1 and Study 2 also both found relationships between integrity and ability indices. Although the observed relationships in both studies were very small, their nature and direction differed diametrically. Study 1 suggested a respondent's crystallised ability was inversely related to how likely they were to misbehave in the workplace. Study 2 reported no significant negative relationship between crystallised ability and integrity scores. The second study instead reported a positive relationship between fluid ability and integrity test indicators of a propensity to engage in counterproductive behaviour. In other words, a respondents fluid ability was directly related to how likely they were to misbehave in the workplace.

Although the directionality and ability index differed across Study 1 and Study 2, both studies considered the relationship observed between integrity and ability to be a function of integrity item and participant characteristics. In Study 1 it was suggested that integrity items may vary for required reading level and those with superior written language comprehension (i.e., crystallised ability) may be more likely to ascertain meaning and answer accordingly. The items in Study 2's integrity scale were written to maximise Greenberg's (1997) geometry of theft rationalisations. This involved providing a hypothetical context in which rationalisation and justification were possible. It was argued that people with relatively higher problem solving capability may more readily discern the likelihood that such rationales are applied by those feeling mistreated, impoverished, desperate, and/or exploited. This recognition could result in more capable problem solvers affirming such items without implying the standard heightened risk of CWBs. The principle of parsimony might suggest an alternative explanation. The small and inconsistent relationships across studies and their lack of consistency with previous suggestions that integrity and ability are unrelated (e.g., Cullen & Sackett, 2004; Ones & Viswesvaran, 1998a; Sackett & Wanek, 1996), could mean that such relationships are a consequence of family-wise statistical errors. Post-doctoral work could explore the truth of observed relationships by examining differential item functioning amongst those of varying ability.

Another interesting disparity in findings between Study 1 and Study 2 concerned the relationship between neuroticism and ability. Study 1 found neuroticism negatively related to ability indices, which is consistent with previous findings (Ackerman & Heggestad, 1997; Austin et al., 2002;

Hembree, 1988; Kyllonen, 1997; Seipp, 1991). Yet Study 2 reported no significant correlations between neuroticism and ability indices. The cause of this difference may reflect the differences in neuroticism scales used within these investigations. Study 2's neuroticism scale comprised items only relating to mercurial and angst facets. As previously noted, Study 1's neuroticism scale comprised additional items relating to a suspicion/cynicism facet. That Study 1's suspicion facet was the neuroticism facet most strongly associated with ability strongly supports the earlier suggestion that it may be the consistent cognitions/schemas associated with neuroticism that are more important to understanding its negative relationship with ability. This may question the previously proffered neural efficiency explanation. This explanation suggested neurotic respondents did less well on ability assessments due to experiencing cognitively taxing levels of anxiety (Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Zeidner & Matthews, 2000). The refutation of this explanation suggests that observed relationships between neuroticism and ability are not artefacts of reactions to testing conditions.

Section 6.2: Adverse Impact Risk and Reduction for those identifying as Māori

This dissertation's second research objective was prompted by concerns that Māori may experience lower hiring ratios than non-Māori when ability assessment outcomes are a primary selection criteria. Aside from reducing diversity, having lower hiring ratios for Māori is inconsistent with the spirit of the State Sector Act (1988) and the principle of partnership within the Treaty of Waitangi (1975).

Study 1 supported previous findings that Māori tended to score lower on average for ability indices than non-Māori within a selection context (Guenole et al., 2003). Yet the significant differences were limited to Gc and *g* indices. This suggests that Māori perform less well than non-Māori on scales related to education, acculturation, and exposure. It also suggests that differences on scales relating to more innate potential are less strong or noteworthy. This was seen as having positive implications for future reductions in the observed disparities. No significant differences were observed in Study 2. However, the small number of Māori participants (N = 22) leaves the robustness of this finding in some doubt.

Study 1 and Study 2 both suggested that no significant or substantial differences were present between Māori and non-Māori for assessments of integrity, conscientiousness, or neuroticism. The parity for integrity scores was consistent with similar findings internationally (Ones, 2003; Sackett & Wanek, 1996). The consistency across conscientiousness scales replicated prior research within the context of workplace testing, but the absence of differences on neuroticism represented a contrary finding within this same research (Packman et al., 2005).

Adverse impact calculations for ability assessments in Study 1 suggested that in situations where only 1% of applicants are hired, the selection ratio for Māori versus non-Māori is likely to be 1:5 (Sackett & Ellingson, 1997). However, this ratio was likely to increase to 3:5 if ability tests were used in conjunction with integrity, conscientiousness, and neuroticism assessments. This represents a substantial increase in hiring ratios for Māori. The lack of significant differences between Māori and non-Māori on scales reported in Study 2 meant adverse impact calculations were not undertaken for the second study.

To reiterate an earlier caveat, the preceding discussion is not intended to resolve issues relating to adverse impact. However, it is hoped that it will contribute to the theoretical and empirical foundation that will allow future assessment and selection practices to more radically reduce adverse impact. The next section focuses upon this dissertation's third research objective.

Section 6.3: Response Time

The third research objective of this dissertation was motivated by two quite distinct considerations. The first was that response time could be used as part of the scoring criteria for ability assessments. This was based upon suggestions that the length of time taken by a respondent to answer ability items provides important information on their ability (Glickman et al., 2005; Scrams & Schnipke, 1997). Such a relationship could have impacted upon the way in which ability assessments are administered and scored. The second consideration was that response time information could help clarify previously observed negative relationships between ability scores and the personality traits of conscientiousness or neuroticism.

The outcomes of this research objective were discussed extensively in Chapter 5. To briefly recap, the relationship between respondent ability and the average length of time taken to complete items or scales did not lend itself to use within a scoring algorithm. The relationships were weak and inconsistent in directionality. The differing cognitive requirement of fluid and crystallised tasks was an explanation consistent with the observed relationships. Again consistent with these differing information-processing requirements, respondents scoring higher on conscientiousness were found to take longer on fluid items, but not crystallised items. Coupled with the fact that higher scorers on fluid ability tended to take longer to answer fluid items, more conscientious respondents taking longer on fluid items may explain the lack of negative relationship between conscientiousness and fluid ability on assessments of moderate difficulty.

Wood and Englert (2009) had thought some lower ability respondents used the adoption of more conscientious behaviour to compensate solely for relatively less knowledge (i.e., crystallised ability). However, the finding more conscientious respondents spend longer on fluid items suggests that the application of that same conscientious behaviour during fluid ability testing may be advantageous enough to mask an otherwise equally negative relationship between conscientiousness and fluid ability.

In another interesting finding only the temper loading neuroticism facet displayed any relationship with the length of time taken to complete ability assessments. This ran counter to expectations that anxiety related tendencies were more responsible for the negative relationship between ability and neuroticism (e.g., Chamorro-Premuzic & Furnham, 2004; Moutafi et al., 2005; Zeidner & Matthews, 2000). The next section of this chapter examines general limitations across both Study 1 and Study 2.

Section 6.4: General Limitations

The general limitations discussed in the following subsections apply to both studies undertaken.

Section 6.4.1:Design

The failure to directly test job performance was a serious limitation. The relatively small relationships amongst predictors in both studies provided tentative support for combining scales of integrity, conscientiousness, neuroticism, and fluid and crystallised intelligence with the aim of increasing predictive validity. However, knowledge of relationships between predictors and job performance was solely dependent on the literature review. Directly investigating the relationships among the predictors tested and the criterion of job performance would have been a superior approach. This would have provided a stronger indication of whether or not the predictors examined were accounting for unique variance in job performance. It would also have enabled an assessment of which combinations of predictors could account for the most variance in job performance.

The failure to gather job performance data also impacted upon adverse impact and criterion validity calculations. The adverse impact reduction formula used in Study 1 assumed all predictors included were of equal importance. Yet most predictive composites are scored on the basis of regression weights proportional to the amount of unique variance explained by component parts. This is the optimal weighting methodology for a composite of predictors in terms of the validity criterion (De Corte et al., 2008). In line with this, the attempt to achieve both increased validity and adverse impact reduction in a composite is a form of multi-objective optimisation (Sackett, De Corte, & Lievens, 2008). Based upon differences in predictive validity reported in the literature review, it is reasonable to suspect that reductions in adverse impact achieved by combining scales may have been smaller had these scales have been weighted on the basis of regression scores.

A further limitation was the use of a cross-sectional instead of a longitudinal approach. This resulted in the inability to determine whether observed results had a transitory age-related existence, were pervasive across age cohorts or the lifespan, or were explicable via a shared developmental trajectory among factors (Bordens & Abbott, 1996; Pedhazur & Schmelkin, 1991).

Section 6.4.2: Generalisability

The capacity to make generalised statements was reduced by use of non-randomly selected participant groups. Confidence in generalisations made from investigative findings depends upon the distribution of participant characteristics being representative of the population to which generalisations are directed. This can be achieved through the use of probability theory in the random selection of individuals from an identified population (Shaughnessy, Zechmeister, & Zechmeister, 2003). Both studies in this dissertation instead used participant groups selected on the basis of convenience. This increases the likelihood that the distribution of characteristics within these participant groups may have systematically differed from the broader New Zealand employee population's distributions, a concern supported by section 5.1.1's detailing of the relative underrepresentation of males and Māori and overrepresentation of higher educated participants. As a result of this limitation generalised statements to any greater population are inappropriate (Pedhazur & Schmelkin, 1991).

The very small number of participants identifying as Māori (N = 46 and N = 22) also greatly limits any attempt at generalising results based on ethnicity related analyses. Conducting analyses on such small groups further increases the potential for non-representative distributions to artificially impact upon results.

Section 6.4.3: Analyses and statistical power

A statistical issue in both studies concerns the hazards of interpreting results (e.g., correlation coefficients) that are significant as an artefact of a large number of tests or large participant group size. On this basis both studies used Bonferroni corrections to avoid inflated family-wise errors (Winer et al., 1991). The high risk of β errors accompanying Bonferroni adjustments was minimised in both studies by generally retaining recommended power (Cohen, 1988; Murphy, Myors, & Wolach, 2009; Spicer, 2005). Ferguson's (2009) recommendations were also followed to address the issue of strength and increase the likelihood that relationships of practical significance would be given the greatest attention.

The use of Sackett and Ellingson's (1997) formula for calculating potential reductions in adverse impact was another analytic limitation. Although the calculated reductions in adverse impact were encouraging, the composite design goal of this investigation encompassed both reductions in adverse impact and likely gains in incremental validity. A better way to balance these potentially conflicting goals may have been via *pareto-optimal calculations*, which present a means of modelling and attempting to balance potential validity losses resulting from attempts to reduce adverse impact (De Corte, Livens, & Sackett, 2007, 2008; Potosky et al., 2005).

Section 6.4.4: Measurement

A measurement limitation concerned the legitimacy of comparing ability tests, designed to measure a maximum performance capability, and personality inventories, designed to measure typical behaviour (Goff & Ackerman, 1992; Most & Zeidner, 1995).

Another measurement limitation across both studies concerned how well the scales used sampled the intended construct domains. The first study used scales for facets not originally designed to provide coverage of the Big Five domains, but instead shown to fit this model *ex post facto*. The second study failed to find factorial justification for any second order Big Five factors. This reduces confidence in the cross-assessment generalisability of results for either study.

A major limitation of the scales employed across both studies concerns their inability to take into account other factors relevant to trait expression. As previously mentioned, research into the situational specificity of personality-job performance relations has identified a variety of situational features that operate at task, social, and organisational levels which are relevant to trait expression (Tett & Burnett, 2003; Tett & Guterman, 2000). The use and design of conscientiousness and neuroticism scales that do not take such factors into consideration is likely to limit the explanatory power of these scales and predictions made on their basis (Cervone, 2005; Leising et al., 2010; Read et al., 2010; Srivastava et al., 2010). The next section will examine potential future directions for subsequent research.

Section 6.5: Future directions

The primary objective of Study 1 and Study 2 was to examine relationships amongst integrity, conscientiousness, neuroticism, and ability. An applied reason for knowledge of such relationships was tentative support for the combined use of these scales in a selection testing context. While observed relationships among these predictors support their combined use, future research must directly examine their combined contribution to the prediction of job performance and associated capacity to reduce adverse impact. This will also facilitate the testing of the hypothesis that the scales chosen will provide coverage of both task and contextual performance.

The studies undertaken in this dissertation used integrity test results as an indirect indicator of counterproductive workplace behavioural propensity. On this basis another important direction for future research is directly testing CWB relationships with scales or neuroticism, conscientiousness, and fluid and crystallised intelligence. This would facilitate much stronger and more justified claims regarding the interactions between these tendencies and abilities.

Future research should also seek to more generally replicate and extend the findings of this dissertation. Limitations concerning the results across these studies were previously discussed and replication is recommended to test the impact such limitations may have had (Cohen et al., 2003; Cook & Campbell, 1979; Shaughnessy et al., 2003).

Section 6.6: Final Thoughts

This dissertation set out to achieve two aims. The first and foremost of these was to investigate the research objectives stated in Chapter 3. The second involved building a new composite assessment of integrity, conscientiousness, neuroticism, and ability. These aims and their success have been addressed elsewhere. This section instead discusses some of the more interesting general themes that have emerged during this dissertation.

One of the most interesting outcomes of exploring relationships among predictor scales was how the level at which analysis was undertaken impacted upon observed relationships. For example, no significant relationship was observed between neuroticism and the length of time taken to complete ability indices. However, when re-examined at the facet level a significant relationship between the length of time taken to answer the fluid ability scale and the neuroticism facet of mercurial was identified. Numerous other examples of relationships between predictors varying at the facet level were reported in Study 1 and Study 2. This theme reinforces the benefit of adopting a fine-grained approach to the analyses of predictor relationships. While the Big Five model is a useful organising device, its traits are so broad that they can obscure relationships among other personality traits and construct domains. The same appears to be true for composite ability scores that do not take into account the very different cognitive processes associated with fluid and crystallised ability.

Another theme to emerge from exploring relationships among predictors was the way in which relationships found differed from previous findings depending upon how traits with similar labels were measured in different assessments. For example, Study 1 found relationships between integrity and the conscientiousness facet of self-control far less prominent than previous research (e.g., Wanek, 1995). However, this was not thought to reflect some novel difference of great pith and moment. It was instead explained as a consequence of the "self-control" facet sampling quite different content across personality assessments. This reinforces the importance of not just relying upon trait labels to understand what is being measured.

Findings within analyses examining the relationship among predictor outcomes and the length of time taken to complete ability scales also revolved around a couple of unifying themes. The first theme concerned the cognitive resources/information-processing involved in completed ability items. The relationship between ability and the length of time taken to complete an ability scale appeared dependent upon the cognitive resources involved in demonstrating the ability in question. For crystallised ability this meant that respondents of higher ability tended to complete the scale relatively quickly, but for the cognitive resource hungry fluid ability items, respondents of higher ability tended to take longer than others to complete the scale. The item-by-item timing procedure employed within this investigation also precluded the claim that these findings reflected potential confounds on the basis of poor pacing strategies. Findings concerning the differing relationships between item difficulty and response times for fluid and crystallised ability further supported this differential cognitive process explanation for findings. These findings suggested that more difficult crystallised items were only answered more quickly by those who knew the answer. For the more cognitively taxing fluid scale, items of more difficulty were answered slower than easier items regardless of response accuracy. The unifying theme of these results concerns the impact that cognitive requirements have on the length of time taken to answer questions.

The second and closely related theme within the response time data concerned the impact personality traits could have on ability scores. Part of the motivation for the third research objective was to test the intelligence compensation theory's argument that negative relationships between ability and conscientiousness reflect a compensation mechanism (Wood & Englert, 2009). It had been hypothesised that this negative relationship was instead a consequence of conscientious respondents disadvantaging themselves through a disproportionate diligence on items presented early in assessments (Jensen, 1998). Study 2 replicated the negative relationship between conscientiousness and ability (Wood & Englert, 2009) despite controlling for this pacing-confound by providing per item time-limits. Somewhat ironically in light of the original motivation, the results of Study 2 also suggested that the absence of a negative relationship between conscientiousness and fluid ability might be due to conscientious respondents taking longer to answer fluid items at the threshold of their maximum ability. In other words, scale outcomes may be affected by both the time items require for information-processing and the trait-related tendencies of respondents. The finding that the mercurial facet of neuroticism was negatively related to response time for fluid ability is also consistent with this theme. It suggests that previously observed negative relationships between ability and neuroticism (e.g., Moutafi et al., 2005) may have less to do with anxiety and more to do with some respondents' increased tendency to give up in frustration or annoyance rather continue with solution seeking behaviour. If any of these relationships really are a consequence of test-taking environments or behaviours, it may be possible to reduce their impact by providing respondents with very explicit and specific information on the nature of the items encountered and appropriate pacing-strategies.

In conclusion, the most consistent theme emerging from this dissertation may be the importance of not deifying scales. Scales are not direct measures of constructs, and constructs provide little more than a glimpse of the complex interactions, influences, and drives associated with human behaviour. Relationships between measures can be obscured based upon levels of analysis undertaken, or confounds associated with interactions between personality tendencies and the administrative process. Increasing the accuracy of scale based inferences requires an understanding of what exactly is being assessed, what the potential compensatory influences might be, and where sources of error may arise. This dissertation has made a small contribution in this regard.

References

Ackerman, P. L., & Goff, M. (1994). Typical intellectual engagement and personality: reply to Rocklin (1994). *Journal of Educational Psychology, 86,* 150-153.

Ackerman, P. L., & Heggestad E., D. (1997). Intelligence, personality and interests: evidence for overlapping traits. *Psychological Bulletin, 121,* 219-245.

Ackerman, P. L., & Humphreys, L. G. (1990). Individual differences theory in industrial and organisational psychology. In M. Dunnette, & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed.), volume 1, (pp.223-283). Palo Alto, CA: Consulting Psychologists Press, Inc.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology,* 1, 876-898.

Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging,* 14, 16-46.

Ackerman, T. A. (1992). *Assessing construct validity using multidimensional item response theory*. Paper presented at the annual meeting of the American Educational Research Association, CA.

Akrami, N., Hedlund, L., & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: the inverted-u effect revisited. *Personality and Individual Differences, 43*, 611-618.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory* (rev. ed.). Long Grove, IL: Waveland Press.

Alicke, M. D., & Largo, E. (1995). The role of the self in the false consensus effect. *Journal of Experimental Social Psychology, 31,* 28-47.

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7,* 32-39.

Allport, G. W. (1937). *Personality: a psychological interpretation.* New York, NY: Henry Holt and Company.

Allport, G. W. (1968). *The person in psychology: selected essays.* Boston, MA: Beacon Press.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47),* (whole No.211), 171.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing* (6th ed). Washington, DC: Author.

Anastasi, A. (1985). The use of personal assessment in industry: Methodological and interpretative problems. In H. J. Bernardin & D. A. Browns (Eds.), *Personality assessment in organizations* (pp. 1-20). New York, NY:Praeger.

Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A cricital review and agenda for future research. *International Journal of Selection and Assessment, 11*, 000-000.

Anderson, N. (2010). Perceived job discrimination (PJD): Toward a model of applicant propensity to case initiation in selection. *International Journal of Selection and Assessment, 17,* 335-345.

Anderson, N., Born, M., & Cunningsnell, N. (2001). Recruitment and selection: applicant perspectives and outcomes. In N. Anderson, D. S., Ones, H. K. Sinangil, and C. Viswesvaran (Eds.), *Handbook of Industrial, Work & Organizational Psychology, Vol. 1: Personnel Psychology* (pp. 200-218). London: Sage.

Aquino, K., Lewis, M. U., & Bradfield, M. (1999). Justice constructs, negative affectivity, and employee deviance: a proposed model and empirical test. *Journal of Organizational Behavior, 20,* 1073-1091.

Arnold, D. W. (1989). *The Reid Report's lack of adverse impact* (Research Memorandum No. 22). Chicagom, IL: Reid Psychological Systems.

Arnold, D., W. (1990). *Legal issues in integrity testing*. Paper presented at the Annual Meeting of the American Psychological Association, Boston, MA.

Aronson, Z. H., & Reilly, R. R. (2006). Personality validity: the role of schemas and motivated reasoning. *International Journal of Selection and Assessment, 14*, 372-380.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49,* 141-168.

Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior, 19,* 289-303.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific factor scales. *Journal of Research in Personality, 29,* 432-442.

Attali, Y. (2005). Reliability of speeded number-right multiple-choice tests. *Applied Psychological Measurement, 29,* 357-368. Retrieved from http://apm.sagepub.com/content/29/5/357

Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., Bent, N., & McInnes, L. (2002). Relationships between ability and personality: does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences*, 32(8), 1391-1411.

Babiak, P., & Hare, R. (2006). *Snakes in suits: when psychopaths go to work.* New York, NY: HaperCollins.

Baccili, P. A. (2001). Organization and manager obligations in a framework of psychological contract development and violation. *Dissertation Abstracts International:* UMI No. 3012291.

Bacha, R. R. (2003). Specifying personality and self-monitoring effects on overall, task, and contextual job performance. *Dissertation Abstracts International:* UMI No. 3098510.

Bagus, K. K. (1988). Industry-specific affirmative action studies. In *Stanton Survey Report.* Charlotte, NC: Stanton Corporation.

Balinsky, B. (1945). The multiple choice group Rorschach test as a means of screening applicants for jobs. *Journal of Psychology, 19,* 203-208.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, N. J.: Prentice-Hall.

Barnabas, B. (1948). Validity of personality and interest tests in selection and placement situations. *Transactions of the Kansas Academy of Science, 51,* 335-339.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Barratt, G. V. (1995). History of personality and intelligence theory and research: The challenge. In D. H. Saklofske & M. Zeidner (Eds.) *International handbook of personality and intelligence*, (pp. 3-13). New York, NY: Plenum Press.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology.* London: Palgrave.

Barrett, P. (2004). The future of ability testing: what happened to the science of mental ability. *Measurement, 2,* 33-36.

Barrett, P., & Kline, P. (1981). The observation to variable ratio in factor analyses. *Journal of Personality and Group Behaviour, 1,* 23-33.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1-25.

Barrick, M. R, Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance, 9,* 9-29.

Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for Retention and Performance. *Human Resource Management, 48,* 183-206.

Bates, T. C., & Eysenck, H. (1998). Intelligence, inspection time, and decision time. *Intelligence, 17,* 523-531.

Bates, T. C., & Rock, A. (2004). Personality and information processing speed: Independent influences on intelligence performance. *Intelligence, 32,* 33-46.

Bates, T. C., & Shieles, A. (2003). Crystallized intelligence as product of speed and drive for experience: the relationship of inspection time and openness to g & gc. *Intelligence*, *31*, 275-287.

Bauer, T. N., Truxillo, D. M., & Paronto, M. E. (2004). The measurement of applicant reactions to selection. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 482-507). Mahwah, NJ: Wiley.

Baumeister, R. F., & Tice, D. M. (1990). anxiety and social exclusion. *Journal of Social and Clinical Psychology, 9,* 165-95.

Bechger, T. M., Huub, M. G., Verstralen, H. F. M., & Beguin, A. A. (2003). Using classical test theory in combination with item response theory. *Applied Psychological Measurement, 27,* 319-334.

Beck, L. R., & St George, R. (1983). The alleged cultural bias of PAT: Reading Comprehension and Reading Vocabulary tests. *New Zealand Journal of Educational Studies, 18,* 32-47.

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85,* 349-360.

Ben-Porath, Y. S., & Waller, N. G. (1992a). “Normal” personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment,* 4, 14-19.

Ben-Porath, Y. S., & Waller, N. G. (1992b). Five big issues in clinical personality assessment: A rejoinder to Costa and McCrae. *Psychological Assessment,* 4, 23-25.

Ben-Shakhar, G., & Bar-Hillel, M. (1993). Misconceptions in Martin and Terris’s (1991) “Predicting infrequent behavior: Clarifying the impact on false positive rates.” *Journal of Applied Psychology, 78,* 148-150.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74,* 137-143.

Berman, J. A. (1993). Validation of the PDI Employment Inventory in a retail chain. *Journal of Business and Psychology, Special Issue: Test Validity Yearbook #2, 7,* 413-420.

Bernardin, H. J., & Cooke, D. K. (1993). Validity of an honesty test in predicting theft among convenience store employees. *Academy of Management Journal, 36,* 1097-1108.

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

Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology, 60,* 271-301.

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology, 62*, 835-863.

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

Bickley, P G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across life span. *Intelligence, 20,* 309-328.

Biderman, M. D., Nguyen, N. T., & Sebren, J. (2008). Time-on-task mediates the conscientiousness-performance relationship. *Personality and Individual Differences, 44,* 887-897.

Billings, S. W. (2001). Clarifications of the relationship between conscientiousness and integrity. *Dissertation Abstracts International:* UMI No. 3037045.

Binet, A., & Simon, T. (1911). New investigations upon the measure of the intellectual level among school children. In H. Gardner, and J. K. Gardner (Eds.). *The development of intelligence in children: The Binet-Simon Scale.* (pp. 274-329). New York, NY: Arno Press.

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

Black, J. (2000). Personality testing and Police selection: utility of the "Big Five". *New Zealand Journal of Psychology, 29,* 2-9.

Blackmore, S. (1999). *The meme machine.* Oxford: Oxford University Press.

Blits, J. H., & Gottfredson, L. S. (1990). Equality or lasting inequality? *Society, 27, 4-8.*

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187-215.

Block, J. (2001). Millennial contrarianism: the five-factor approach to personality description 5 years later. *Journal of Research in Personality, 35,* 98-107.

Bobko, P., & Roth, P. L. (2004). Personnel selection with top-score-referenced bandings: On the inappropriateness of current procedures. *International Journal of Selection and Assessment, 12,* 291-298.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods, 10,* 689-709. Retrieved from http://orm.sagepub.com/cgi/content/abstract/10/4/689-709.

Boehm, V. R. (1972). Negro-White differences in validity of employment and training selection procedures: Summary of recent evidence. *Journal of Applied Psychology, 56,* 33-39.

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic,* June, 35-37.

Bordens, K. S., & Abbott, B. B. (1996). *Research design and methods: A process approach (3rd ed.).* Palo Alto, CA: Mayfield Publishing Company.

Borman, W. C., Hanson, M., & Hedge, J. (1997). Personnel selection. *Annual Review of Psychology, 48,* 299-337.

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

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of rates task performance and interpersonal factors on supervisory and peer performance ratings. *Journal of Applied Psychology, 80,* 168-177.

Borofsky, G. L. (1992). Assessing the likelihood of reliable workplace behaviors: Further contributions to the validation of the Employee Reliability Inventory. *Psychological Reports, 70,* 563-592.

Borofsky, G. L., & Smith, M. (1993). Reductions in turnover, accidents and absenteeism: The contribution of a pre-employment screening inventory. *Journal of Clinical Psychology, 49,* 109-116.

Borsboom, D. (2006a). The attack of the psychometricians. *Psychometrika, 71,* 425-440.

Borsboom, D. (2006b). Can we bring about a velvet revolution in psychological measurement? A rejoinder to commentaries. *Psychometrika, 71,* 463-467.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist,* 44, 576-578.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process.* Chicago, IL: University of Chicago Press.

Boyle, G. J. (1989). Re-examination of the major personality type factors in the Cattell, Comrey, and Eysenck scales: Were the factor solutions by Noller et al. optimal? *Personality and Individual Differences,* 10, 1289-1299.

Boyle, G. J. (1991). Experimental psychology does require a multivariate perspective. *Contemporary Psychology,* 36, 350-351.

Boyle, G. J., & Cattell, R. B. (1984). Proof of situational sensitivity of mood states and dynamic traits – ergs and sentiments – to disturbing stimuli. *Personality and Individual Differences,* 5, 541-548.

Boyle, G. J., Stankov, L., & Cattell, R. B. (1995). Measurement and statistical models in the study of personality and intelligence. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, (pp.417-446). New York, NY: Plenum Press.

Brand, C. (1994). Open to experience – closed to intelligence: Why the 'big five' are really the 'comprehensive six'. *European Journal of Personality,* 8, 299 – 310.

Brand, C. R. (1996). *The g factor.* Chichester: Wiley.

Bridgeman, B. (1980). Generality of a "fast" or "slow" test-taking style across a variety of cognitive tasks. *Journal of Educational Measurement, 17,* 211-217. Retrieved from http://www.jstor.org/stable/1434836.

Birnbaum , A. (1968), Some Latent Trait Models and Their Use in Inferring an Examinee's Ability, In F.M Lord & M.R. Novick (Eds.), *Statistical Theories of Mental Test Scores*, (pp. 397-472), Reading, MA: Addison-Wesley.

Bjorner, J.B., Smith, K.J., Stone, C, & Sun, X.W. (2007). *IRTFIT: A macro for item fit and local dependence tests under IRT models* [Computer program]. QualityMetric, Inc.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431–444.

Brown, G. S., & White, K. G. (2009). Measuring discriminability when there are multiple sources of bias. *Behavior Research Methods,* 41, 75-84.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111-150.

Budd, R. (1993). *Manual for the technical test battery.* Letchworth, Herts: Psytech International Ltd.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Psychology, 19,* 176-199.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York, NY: Wiley.

Buss, D. M. (1991). Evolutionary personality psychology. *Annual Review of Psychology, 42,* 459-91.

Byle, K. A., & Holtgraves, T. M. (2008). Integrity testing, personality, and design: interpreting the personnel reaction blank. *Journal of Business Psychology, 22,* 287-295.

Byrne, R. W., & Whiten, A. (Eds.) (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in Monkeys, Apes, and Humans.* Oxford: Oxford University Press.

Byrne, Z. S., Stoner, J., Thompson, K. R., & Hochwarter, W. (2005). The interactive effects of conscientiousness, work effort, and psychological climate on job performance. *Journal of Vocational Behavior, 66,* 326-338.

Camara, W. J., & Schneider, D. L. (1994). integrity tests: Facts and unresolved issues. *American Psychologist, 49,* 112-119.

Cameron, L. C., Ittenback, R. F., McGrew, K. S., Harrison, P. L., Taylor, L. R., & Hwang, Y. R. (1997). Confirmatory factor analysis of the K-ABC with gifted referrals. *Educational and Psychological Measurement, 57,* 823-840.

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

Campbell, J. P. (1996). Group differences and personnel decisions: validity, fairness, and affirmative action. *Journal of Vocational Behavior, 49,* 122-158.

Campbell, J. P. (1999). The definition and measurement of performance in the new age. In D. R. Ilgen, & E. D. Pulakos, E. D. (Eds.), *The changing nature of performance: Implications for staffing, motivation, and development* (pp. 399-431). San Francisco: Jossey-Bass.

Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organisations* (pp. 258-299). San Francisco: Jossey-Bass.

Campion, M. A., Fink, A. A., Ruggeburg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practice in competency modelling. *Personnel Psychology, 64*, 225-263.

Campion, M. A., Outtz, J. L., Zedeck, S., Schmidt, F. L., Kehoe, J. F., Murphy, K. R., & Guion, R. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology, 54,* 149-185.

Caron, S. J. (2003). Personality characteristics related to counterproductive behaviors in the workplace. *Dissertation Abstracts International:* UMI No. 1412241.

Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed). *Handbook of human intelligence, (pp. 3-31).* New York, NY: Cambridge University Press.

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 1-33). New York, NY: Academic Press.

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

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

Cascio, W. F. (1991). *Applied psychology in personnel management* (4th ed). Englewood Cliffs, NJ: Prentice Hall.

Cascio, W. F., Jacobs, R., & Silva, J. (2010). Validity, utility, and adverse impact: Practical applications from 30 years of data. In J. L. Outtz (Ed.), *Adverse impact: implications for organizational staffing and high stakes selection* (pp. 371-289). New York, NY: Routledge.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1943a). The description of personality: Basic traits resolved into clusters. *Journal of Personality and Social Psychology,* 38, 476-506.

Cattell, R. B. (1943b). The measurement of adult intelligence. *Psychological Bulletin,* 40, 153-193.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York, NY: World Book Company.

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

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Cattell, R. B. (1978). *The scientific use oof factor analysis in the behavioural and life sciences.* New York, NY: Plenum Press.

Cattell, R. B. (1979). Are culture fair intelligence tests possible and necessary? *Journal of Research and Development in Education,* 12, 3-13.

Cattell, R. B. (1982). *The inheritance of personality and ability: Research methods and findings*. San Diego: Academic Press.

Cattell, R. B. (1987c). The discovery of fluid and crystallized general intelligence. In R, B. Cattell (Ed), *Intelligence: Its structure, growth and action,* (pp. 87-121). New York, NY: North-Holland.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the sixteen-personality factor questionnaire (16PF).* Champaign, IL: Institute for Personality and Ability testing.

Cervone, D. (2005). Personality architecture: within-person structures and processes. *Annual Review of Psychology, 56,* 423-452.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37,* 319-338.

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

Chan, D. (2005). Current directions in personnel selection research. *Current Directions in Psychological Science, 14,* 220-223.

Chang, H.H., & Ying, Z. (2008). To weight or not to weight? Balancing influence of initial items in adaptive testing. *Psychometrika,* 73, 441-450.

Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment, 11*, 113-120.

Chen, P. Y., & Spector, P. E. (1992). Relationships of work stressors with aggression, withdrawal, theft and substance use:Anexploratory study. *Journal of Occupational and Organizational Psychology*, *65***,** 177–184.

Chernyshenko, S., Stark, S., & Drasgow, F. (2008). *Adaptive General Reasoning Test (AdaptGRT): Report to Psychometrics Ltd.* Urbana, IL: Drasgow Consulting Group.

Cherrington, D. (1989). *Validation of the Applicant Review: Analysis of adverse impact* (Research Report). Orem, UT: CHC Forecast.

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continumm.

Ciancutti, A., & Steding, T. L. (2001). *Built on trust: Gaining competitive advantage in any organisation.* Chicago, IL: Contemporary Books.

Clark, L. A. (2006). When a psychometric advance falls in the forest. *Psychometrika, 71,* 447-450.

Cleary, T. A., Humphreys, L. G., Kendrick, A. S., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30,* 15-41.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Mahwah, N.J: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Cohen, J., Cohen, C., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cohen, J., & Swerdlik, M. E. (2002). *Psychological testing and assessment*: *An introduction to tests and measurement* (5th ed.). Boston, MA: McGraw Hill.

Cohen, J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). Boston, MA: McGraw Hill.

Comrey, A. L (1988). Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology. *Journal of Consulting and Clinical Psychology,* 56, 754–761.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation, 10.* Retrieved from http://pareonline.net/getvn.asp?v=10&n=7 .

Coyne, I., & Bartram, D. (2002). Assessing the effectiveness of integrity tests: A review. *International Journal of Testing, 2,* 15-34.

Collins, J. M., & Griffin, R. W. (1998). The psychology of counterproductive job performance. In R. W. Griffin, A. O'Leary-Kelly, & J. M. Collins (Eds.), *Dysfunctional behavior in organizations: Violent and deviant behavior,* (pp. 219-242). New York, NY: JAI Press.

Collins, J. M., & Schmidt, F. L. (1993). Integrity, personality, and white-collar crime: A construct validity study. *Personnel Psychology, 46,* 295-311.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences, 34,* 33-39.

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

Cone, J. D., & Foster, S. L. (2001). *Dissertations and theses from start to finish: Psychology and related fields.* Washington, DC: American Psychological Association.

Connelly, B. S., Lilienfeld, S. O., & Schmeelk, K. M. (2006). Integrity tests and morality: Associations with ego development, moral reasoning, and psychopathic personality. *International Journal of Selection and Assessment, 14,* 82-87.

Connerley, M. L., Arvey, R. D., & Bernardy, C. J. (2001). Criminal background checks for prospective and current employees: Current practices among municipal agencies. *Public Personnel Management, 30,* 173-183.

Connerley, M. L., Arvey, R. D., Gilliland, S. W., Mael, F. A., Paetzold, R. L., & Sackett, P. R. (2001). Selection in the workplace: Whose rights prevail? *Employee Responsibilities and Rights Journal, 13,* 1-13.

Connolly, J. J., & Viswesvaran, C. (2000). The role of affectivity in job satisfaction: a meta-analysis. *Personality and Individual Differences, 29,* 265-281.

Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: implications for selection involving multiple predictors. *International Journal of Selection and Assessment, 17*, 47-60.

Cook, M. (2004). *Personnel selection: Adding value through people* (4th ed.). West Sussex, England: Wiley.

Corsini, R. J., & Auerbach, A. J. (1998). *Concise encyclopedia of psychology* (2nd ed). New York: Wiley.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: A primer.* Retrieved from http://www.psych.ucsb.edu/research/cep/primer.html.

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Steinberg & J. C. Kaufman (Eds.), *The evolution of intelligence* (pp. 145-198). Mahwah, NJ: Erlbaum.

Costa, P. T., Fozard, J. L., McCrae, R. R., & Bosśe, R. (1976). Relations of age and personality dimensions to cognitive ability factors. *Journal of Gerontology, 31,* 663-669.

Costa, P. T., Herbst, J. H., & McCrae, R. R. (2002). The replicability and utility of three personality types. *European Journal of Personality*, 16, 573-87.

Costa, P. T., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self reports and spouse ratings on the NEO personality inventory. *Journal of Personality and Social Psychology,* 54, 853-863.

Costa, P. T., McCrae, R. R., Zonderman, A. B., Barbano, H. E., Lebowitz, B, & Larson, D. M. (1986). Cross-sectional studies of personality in a national sample: 2 stability in neuroticism, extraversion, and openness. *Psychology and Aging,* 1, 144-149.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology,* 81, 322-331.

Cowan, N. (2005). Understanding intelligence: A summary and an adjustable-attention hypothesis. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence,* (pp. 469-488). Thousand Oaks: Sage.

Cramer, P. (1999). Ego functions and ego development: defense mechanisms and intelligence as predictors of ego level. *Journal of Personality, 67,* 735-760.

Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology, 80,* 532-537.

Crigmile, P. F., Peruggia, M., & van Zandt, T. (2010). Hierarchical bayes models for response time data. *Psychometrika, 75,* 613-632.

Cronbach, L. J. (1970). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.), (pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed). New York, NY: Harper & Row.

Cropanzano, R., James, K., & Konovsky, M. A. (1993). Dispositional affectivity as a predictor of work attitudes and job performance. *Journal of Organizational Behavior, 14,* 595-606.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York, NY: Wiley.

Cullen, M. J., & Sackett, P. R. (2004). Integrity testing in the workplace. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 149-165). Mahwah, NJ: Wiley.

Cunningham, M. R. (1989). Test-taking motivation and outcomes on a standardised measure of on-the-job integrity. *Journal of Business and Psychology, 4,* 119-127.

Cunningham, M. R., Wong, D. T., & Barbee, A P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology, 79,* 643-658.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology, 90,* 1241-1255.

Dalal, R. S., Lam, H., Weiss, H. M., Welch, E. R., & Hulin, C. (2009). A within-person approach to work behaviour and performance: concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance. *Academy of Management Journal, 52,* 1051-1066.

Daniel, R. C., & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement, 34,* 348-364.

Danthiir, V., Roberts, R. D., Pallier, G., & Stankov, L. (2001). What the nose knows: Olfaction and cognitive abilities. *Intelligence, 29,* 337-361.

Darwin, C. R. (1859). *The origin of species by means of natural selection.* London: Encyclopaedia Britannica, Inc.

Dawkins, R. (1986). *The blind watchmaker.* Harlow: Longman.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* New York, NY: Oxford University Press.

Deary, I. J. (2001). *Intelligence, a very short introduction.* New York, NY: Oxford University Press.

De Corte, W., Lievens, F., & Sackett, P. R. (2006). Predicting adverse impact and mean criterion performance in multistage selection. *Journal of Applied Psychology, 91*, 523-537.

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combing predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology, 92,* 1380-1393.

De Corte, W., Lievens, F., & Sackett, P. R. (2008). Validity and adverse impact potential of predictor composite formation. *International Journal of Selection and Assessment, 16,* 183-194.

DeCoster, J. (1998). Overview of Factor Analysis. Retrieved from http://www.stat-help.com/notes.html

de la Haye, A. (2000). A methodological note about the measurement of the false consensus effect. *European Journal of Social Psychology, 30,* 569-581.

de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size: a higher order IRT model approach. *Applied Psychological Measurement, 34,* 267-285. Retrieved from http://apm.sagepub.com/content/34/4/267.

Demetriou, A., & Papadopoulos, T. C. (2004). Human intelligence: From local models to universal theory. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 445-474). Cambridge, U.K.: Cambridge University Press.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York, NY: Simon & Schuster.

De Raad, B., et al. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology, 98*, 160-173.

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138–1151.

DeYoung, C.G., Peterson, J.B., & Higgins, D.M. (2001). Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual Differences 33*, 533–552.

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

Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology, 73,* 1246-1256.

Digman, J. M, & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research,* 16, 149-170.

Dirks, K. T., & Daniel P. S. (2004). Trust in leaders: Existing research and emerging issues. In R. M. Kramer, & K. S. Cook (Eds.), *Trust and distrust in organizations: Dilemmas and approaches* (pp. 21-41). New York: Sage.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology, 92,* 616-627.

Dilchert, S., Ones, D. S., Viswesvaran, C., Deller, J. (2006). Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 48*, 209-225.

Dillon, C. K., & Harris, W. G. (1988). *Investigating counterproductivity and inventory shrinkage with an integrity measure. Technical Report.* Charlotte, IL: Stanton.

Dion, K. L., & Yee, P. H. N. (1987). Ethnicity and personality in a Canadian context. *Journal of Social Psychology, 127,* 175-182.

Dixon, M. E., & Johnston, R (1980). *The roots of individuality.* Belmont, CA: Brooks-Cole.

Donnellan, M. B., Oswald, F. L., & Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18*, 192-203.

Drasgow, F. (2003). Intelligence and the workplace. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 107-130). New York, NY: Wiley.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19*, 143-165.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16,* 939-944.

Dunleavy, E., & Gutman, A. (2009). Fasten your seatbelts: Supreme Court to hear Ricci v. Destefano. *The Industrial-Organizational Psychologist, 46,* 31-43.

Dunn, P.A. (1995). Pre-employment Referencing Aids your Bottom-line. *Personnel Journal*, *74*, 68-75.

Dunn, W. S., Mount, M. K., Barrick, M. R., Ones, D. S. (1995). Relative importance of personality and general mental ability in managers' judgments of applicant qualifications. *Journal of Applied Psychology, 80,* 500-519.

Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. *Educational and Psychological Measurement, 57,* 937-948.

Eatwell, J. (1997). Using norms in the interpretations of test results. In H. Love & W. Whittaker (Eds). *Practice issues for clinical and applied psychologists in New Zealand* (pp. 268-276). Wellington: The New Zealand Psychological Society.

Ebel, R.L., & Frisbie, D.A. (1986).*Essentials of educational measurement.* Englewood Cliffs, N.J: Prentice-Hall.

Egan, V. (1989). Links between personality and, ability and attitudes in a low-IQ sample. *Personality and Individual Differences, 10*, 997-1001.

Ellingson, J. E., Sackett, P R., Connelly, B. S. (2007). Personality assessment across selection and development contexts: insights into response distortion. *Journal of Applied Psychology, 92*, 386-395.

Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*, 122-133.

Ellis, A., & Conrad, H. S. (1948). The validity of personality inventories in military practice. *Psychological Bulletin, 45,* 385-426.

Elovainio, M., Kivimaki, M., Vahtera, J., Virtanen, M., & Keltikangas-Jarvinen, L. (2003). Personality as a moderator in the relations between perceptions of organizational justice and sickness absence. *Journal of Vocational Behavior, 63,* 379-395.

Embretson, S. E. (1983). Construct validity: construct representation versus nomothetic span. *Pscyhological Bulletin, 93,* 179-197.

Embretson, S. E. (1985). *Test design: Developments in psychology and psychometrics.* Orlando, FL: Academic Press.

Embretson, S. E., & Reise, R. M. (2000). *Item Response Theory for Psychologists.* New York, NY: Plenum.

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

Entink, R. H. K., Fox, J-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modelling of accuracy and speed of test takers. *Psychometrika, 74,* 21-48.

Everitt, B.S. (2002). *The Cambridge Dictionary of Statistics* (2nd ed.). London: Cambridge University Press.

Evers, A., Anderson, N., & Voskuijl, O. (2005), *The Blackwell handbook of personnel selection*. Malden, MA: Blackwell Publishing.

Evers, A., Te Nijenhuis, J., & van der Flier, H. (2005). Ethnic bias and fairness in personnel selection: evidence and consequences. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection,* (pp. 306-328). Malden, MA: Blackwell Publishing.

Ewen, R. B. (1988). *An introduction to theories of personality* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Ewen, R. B. (1998). *Personality: A topical approach: Theories, research, major controversies, and emerging findings*. Mahwah, NJ: Lawrence Erlbaum.

Ewen, R. B. (2003). *An introduction to theories of personality* (6th ed.). Mahwah, N.J: Lawrence Erlbaum.

Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London: Methuen.

Eysenck, H. J. (1971). Relation between intelligence and personality. *Perceptual and Motor Skills, 32*, 637-638.

Eysenck, H. J. (1992). Four ways five factors are *not* basic. *Personality and Individual Differences,* 6, 667-673.

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

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.

Feist, G. J., & Barron, F. X. (2003). Predicting creativity from early to late adulthood: Intellect, potential, and personality. *Journal of Research and Personality,* 37, 62-88.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 523-538.

Fergusson, D. M., Horwood, L. J., & Swain-Campbell, N. (2003). Ethnicity and Criminal Convictions: Results of a 21-year Longitudinal Study. *Australian & New Zealand Journal of Criminology,* 36, 354-409. Retrieved from http://anj.sagepub.com/content/36/3/354.1.

Ferrando, P. J., & Anguiano-Carrasco, C. (2009). Assessing the impact of faking on binary personality measures: an IRT-based multiple-group factor analytic procedure. *Multivariate Behavioral Research, 44,* 497-524.

Ferrando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement, 61,* 997-1012.

Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement, 31,* 525-543. Retrieved from http://apm.sagepub.com/content/31/525

Field, A., (2005). *Discovering Statistics Using SPSS* (2nd ed.). London: Sage.

Fifield, J.K., & Donnell, A.A. (1980). *Socioeconomic status, race, and offending in New Zealand,* (Research Report No 6). Wellington, New Zealand: Government Printer.

Fine, S., Horowitz, I., Weigler, H., & Basis, L. (2010). Is good character good enough? The effects of situational variables on the relationship between integrity and counterproductive work behaviors. *Human Resource Management Review, 20*, 73-84.

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11,* 58-66.

Flanagan, D. P., & McGrew, K. S. (1995). The field of intellectual assessment: A current perspective. *The School Psychologist, 49,* 7-14.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & R. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 314-325). New York, NY: Guilford Press.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary *Gf-Gc* theory: Joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology, 36,* 151-182.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Floyd, R. G., McGrew, K. S., Barry, A. Rafael, F., & Rogers, J. (2009). General and Specific Effects on Cattell–Horn–Carroll Broad Ability Composites: Analysis of the Woodcock–Johnson III Normative Update Cattell–Horn–Carroll Factor Clusters Across Development. *School Psychology Review, 38,* 249-265.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J. A. (2001). *The mind doesn't work that way.* Cambridge, MA: MIT Press.

Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61,* 579-616.

Forlano, G., & Kirkpatrick, F. H. (1945). Intelligence and adjustment measurements in the selection of radio tube mounters. *Journal of Applied Psychology, 29,* 257-261.

Fox, S., & Spector, P. E. (1999). A model of work frustration-aggression. *Journal of Organizational Behavior, 20***,** 915–931.

Fox, S., Spector, P. E., & Miles, D. (2001). Counterproductive work behavior (CWB) in response to job stressors and organizational justice: some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior, 59,* 291-309.

Franz, S. I. (1919). *Handbook of mental examination methods* (2nd ed.). New York, NY: Macmillian.

Frearson W.M., & Eysenck, H.J. (1986). Intelligence, reaction time (RT) and a new "odd-man-out" RT paradigm. *Personality and Individual Differences, 7,* 807-817

Freud, A. (1937). *The ego and the mechanisms of defense*. London: Hogarth.

Freud, S. (1917). *A difficulty in the path of psychoanalysis.* London: Hogarth.

Fry, P. S. (Ed.) (1984). *Changing conceptions of intelligence and intellectual functioning: Current theory and research.* Amsterdam: Elsevier Science.

Furnham, A. (2001). Personality and individual differences in the workplace: Person – organisation –outcome fit. In R. W. Roberts, & R. Hogan (Eds.), *Personality psychology in the workplace,* (pp. 223-251). Washington, DC: American Psychological Association.

Furnham, A., Crump. J., & Whelan, J. (1997). Personality and job satisfaction. *Personality and Individual Differences, 22,* 669-675.

Furnham, A., Forde, L., & Cotter, T. (1998a). Personality and intelligence. *Personality and Individual Differences, 14,* 187-192.

Furnham, A., Forde, L., & Cotter, T. (1998b). Personality scores and test taking style. *Personality and Individual Differences, 24,* 19-23.

Gael, S., & Grant, D. L. (1972). Employment test validation for minority and nonminority telephone company service representatives. *Journal of Applied Psychology, 56,* 135-139.

Galton, F. (1869). Classification of men according to their natural gifts. In J. J. Jenkins, & D. G. Patterson (Eds.), *Studies in individual differences: The search for intelligence.* (pp. 1-17). London: Methuen.

Galton, F. (1885). On the anthropometric laboratory at the late International Health Exhibition. *Journal of the Anthropological Institute, 14,* 205-219.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York, NY: Basic Books.

Geary, D. C., & Huffman, K. J. (2002). Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin, 128,* 667-698.

George, J. M. (1996). Trait and state affect. In K. R. Murphy (Ed.). *Individual differences and behaviour in organizations* (pp. 145-175). San Francisco: Jossey-Bass.

Germain, S., Valois, P., & Abdous, B. (2007). *Eirt – Item Response Theory assistant for Excel.* Retrieved from http://libirt.sourceforge.net/eirt-en/index.html**.**

Ghiselli, E. E., Campbell, J. P., Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: Freeman.

Giacalone, R. A., & Greenberg, J. (1997). *Antisocial behaviour in organizations.* Thousand Oaks, CA: Sage.

Gierl, M. J. Leighton, J. P, & Hunka, S. M. (2005). Exploring the logic of Tatsuoka's Rule-Space Model for test development and analysis. *Educational Measurement: Issues and Practice, 19*, 34-44.

Gies, R. L., & Moon, T. H. (1981). Machiavellianism and deception. *Journal of Personality and Social Psychology, 41,* 403-416.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organisational justice perspective. *Academy of management Review, 18,* 694-734.

Girotto, V., & Tentori, K. (2008). Is domain-general thinking a domain-specific adaptation? *Mind and Society, 2,* 167-175.

Glickman, M. E., Gray, J. R., & Morales, C. J. (2005). Combining speed and accuracy to assess error-free cognitive processes. *Psychometrika, 70,* 405-425.

Glickman, M. E., Seal, P., & Eisen, S. V. (2009). *A non-parametric Bayesian diagnostic for detecting differential item functioning in IRT models*. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=21808376

Goff, M., & Ackerman, P. L. (1992). Personality – intelligence relations: assessment of typical intellectual engagement. *Journal of Educational Psychology*, 84, 537-552.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216-1229.

Goldberg, L. R., Grenier, J. R., Guion, R. M., Sechrest, L. B., & Wing, H. (1991). *Questionnaires used in the prediction of trustworthiness in preemployment selection decisions.* Washington, DC: American Psychological Association.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84-96.

Goldberg, L. R., Sweeney, D., Merenda, P. F., & Hughes, J. E. (1998). Demographic variables and personality: The effects of gender, age, and ethnic/racial status on self-descriptions of personality attributes. *Personality and Individual Differences, 24,* 393-403.

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

Gorman, C. A., Meriac, J. P., Overstreet, B. L., Apodaca, S., McIntyre, A. L., Park, P., & Godbey, J. N. (2011). A meta-analysis of the regulatory focus nomological network: Work-related antecedents and consequences. *Journal of Vocational Behaviour, 90*, 148-161.

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

Gorsuch, R. L., & Cattell, R. B. (1977). Personality and socio-ethical values: The structure of self and superego. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory* (pp. 675-708). New York: Wiley.

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

Gough, H. G. (1972). *Manual for the Personnel Reaction Blank.* Palo Alto, CA: Consulting Psychologists Press.

Gould, S. J. (1993). *The mismeasure of man.* New York, NY: Norton.

Green, A. P. (2003). Identifying the dimensions of integrity: A confirmatory and discriminant validity analysis. *Dissertation Abstracts International:* UMI No. 3113049.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: a cautionary tale. *Psychometrika, 74,* 121-135.

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequality: The hidden cost of pay cuts. *Journal of Applied Psychology, 75,* 561-568.

Greenberg, J. (1998). The cognitive geometry of employee theft: Negotiating “The Line” between taking and stealing. In R. W. Griffin, A. O’Leary-Kelly, & J. M. Collins (Eds.), *Dysfunctional behavior in organizations: Violent and deviant behavior* (pp. 147-193). New York, NY: JAI Press.

Greenberg, J. (2003). *Organizational behaviour: The state of the science* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Greenwald, A. G, McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology,* 74, 1464-1480.

Grimm, P. F., Kohlberg, L., White, S. H. (1968). Some relationships between conscience and attentional processes. *Journal of Personality and Social Psychology, 8,* 239-252.

Grubb, W. L., Whetzel, D. L., & McDaniel, M. A. (2004). General mental ability tests in industry. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 7-21). Mahwah, N.J: Wiley.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence, 29,* 523-535.

Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behaviour. *International Journal of Selection and Assessment, 11,* 30-42.

Guastello, S. J., & Rieke, M. L. (1991). A review and critique of honesty test research. *Behavioral Sciences and the Law, 9,* 501-523.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds), *Handbook of qualitative research,* (pp. 105-117). Boston, M.A.: Sage.

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,* 49-54.

Guilford, J. P. (1956). *Psychometric methods.* New York, NY: McGraw-Hill.

Guilford, J. P. (1959). *Personality.* New York, NY: McGraw-Hill.

Guilford, J. P. (1966). Intelligence: 1965 model. *American Psychologist, 21, 20-26.*

Guilford, J. P. (1967). *The nature of human intelligence.* New York, NY: McGraw-Hill.

Guilford, J. P. (1982). Cognitive psychology’s ambiguities: Some suggested remedies. *Psychological Review, 89,* 48-59.

Guilford, J. P., & Merrifield, P. H. (1960). The structure of intellect model: Its uses and implications. *Reports from the psychological laboratory.* L.A., CA: University of Southern California.

Guion, R. M. (1991). Personnel assessment, selection, and placement. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol.2), (pp. 327-398) Palo Alto, CA: Consulting Psychologists Press.

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

Gustafson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.

Gustafson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35-71). New York, NY: Academic Press.

Gustafsson, J. E. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407-434.

Gutpa, V. (2000). Regression explained in simple terms. Retrieved from http://medresearchconsult.com/regression_explained.pdf.

Hakstian, R. A., & Cattell, R. B. (1974). The checking of primary ability structure on a broader basis of performance. *British Journal of Educational Psychology,* 44, 140-154.

Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*, 309-334.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices, 12*, 38-47.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Hare, R. (1993). *Without conscience: The disturbing world of the psychopaths amongst us*. New York, NY: Simon & Shuster.

Harker, R. K. (1978). Achievement and ethnicity: Environmental deprivation or cultural difference? *New Zealand Journal of Educational Studies, 13,* 107-124.

Harnqvist, K., Gustafsson, J. E., Muthen, B. O., & Nelson, G. (1994). Hierarchical models of ability at individual and class levels. *Intelligence, 18,* 165-187.

Harrell, T. W. (1949). Humm-Wadsworth Temperament Scale and ratings of salesmen. *Personnel Psychology, 2,* 451-495.

Harris, M. M., Dworkin, J. B., Park, J. (1990). Pre-employment screening procedures: How human resource managers perceive them. *Journal of Business and Psychology, 4,* 279-292.

Harris, W. G. (1987). A components analysis of a pre-employment integrity measure: A replicated study. *Psychological Reports, 60,* 1051-1055.

Harris, W. G., & Gentry, M. U. (1992). *Stanton Survey: New Edition Test Manual*. Charlotte, IL: Stanton.

Hartlage, S., Alloy, L. B., Vasquez, C., & Dykman, B. (1993). Automatic and effortful processing in depression. *Psychological Bulletin, 113,* 247-278.

Harvey, R. J. & Hammer, A. L. (1999). Item response theory. *The Counselling Psychologist, 27,* 353-383.

Hawk, J. (1986). Real world implications of *g*. *Journal of Vocational Behavior, 29,* 411-414.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38,* 1128-1142.

Hegarty, W. H., & Sims, J. P. (1978). Some determinants of unethical behavior: An experiment. *Journal of Applied Psychology, 63,* 451-457.

Heim, A. H. (1970). *Intelligence and personality.* Harmondsworth, Middlesex: Penguin.

Heim, A. H., Watt, K. P., & Simmonds, V. (1974). *AH2/AH3 group tests of general reasoning: manual.* Nelson, UK: NFER.

Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research,* 58, 47-77.

Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R.L. Thorndike (Ed.), *Educational Measurement* (p.141). Washington DC: American Council on Education

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: The Free press.

Hershcovis, M. S., et al. (2007). Predicting workplace aggression: a meta-analysis. *Journal of Applied Psychology, 92,* 228-238.

Hesketh, B. (1993). Measurement issues in industrial and organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.). *International review of industrial and organizational psychology.* Vol.8 (pp. 133-172). London: Wiley.

Higgins, E. T., & Scholer, A. A. (2008). When is personality revealed? A motivated cognition approach. In O. P. John, R. W. Robins, and L. A. Pervin (Eds.), *Handbook of personality psychology* (3rd ed.) (pp. 182-207). New York, NY: The Guillford Press.

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology, 55,* 363-396.

Hilliard, P. A. (2001). Comparison of the predictive validity of a written test, an integrity test, a conscientiousness questionnaire, a structured behavioural interview and a personality inventory in the assessment of job applicants' background investigations, and subsequent task and contextual job performance. *Dissertation Abstracts International: Section B: The Sciences &Engineering, 62(6-B), 2981.* (UMI No. 0419-4217).

Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Boca Raton, FL: Chapman & Hall/CRC.

Hogan, J., Barrett, P., Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92,* 1270-1285.

Hogan, J., & Brinkmeyer, K. (1997). Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology, 50,* 587-599.

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology, 74,* 273-279.

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 849-866). New York, NY: Academic Press.

Hogan, R., & DeSoto, C. B., & Solano, C. (1977). Traits, tests, and personality research. *American Psychologist, 32,* 255-264.

Hogan, R., Hogan, J., and Barrett, P. T. (2008) Good Judgment: the intersection of intelligence and personality. In O. Y. Chebykin, G. Z. Bedney, & W. Karwowski (Eds.) *Ergonomics and Psychology: Developments in Theory and Practice*, chapter 17, pp. 357-376. London: CRC Press, Taylor and Francis.

Hogan, R., & Nicholson, R. (1988). The meaning of personality test scores. *American Psychologist, 43,* 621-626.

Hogan, R., & Roberts, B. W. (1996). Issues and nonissues in the fidelity/bandwidth tradeoff. *Journal of Organizational Behavior, 17,* 627-637.

Hogan, R., Rybicki, S. L., Motowidlo, S. J., & Borman, W. C. (1998). Relations between contextual performance, personality, and occupational advancement. *Human Performance, 11,* 189-207.

Holden, R. R., & Kroner, D. G. (1992). Relative Efficacy of Differential Response Latencies for detecting faking on a Self-Report Measure of Psychopathology. *Psychological Assessment, 4*, 170–173.

Holland, D. C., Dollinger, S. J., Holland, C. J., & MacDonald, D. A. (1995). The relationship between psychometric intelligence and the five-factor model of personality in a rehabilitative sample. *Journal of Clinical Psychology,* 51, 79-88.

Holland, P. W., Hoskens, M. (2003). Classical test theory as a first order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika, 68,* 123-149.

Homer. (1974). The Iliad (translated by R. Fitzgerald). In R. Fitzgerald, *Homer The Iliad.* Oxford: Oxford University Press.

Horn, J. L. (1968). Organisation of abilities and the development of intelligence. *Psychological Review,* 75, 242-259.

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. *Annual Review of Psychology*, 27, 437-85.

Horn, J. L. (1977). Personality and ability theory. In R. B. Cattell & R. M. Dreger. *Handbook of modern personality theory,* (pp. 139-166). Washington, DC: A Halsted Press Book

Horn, J. L. (1985). Remodeling old models of intelligence. In B. Wollman (Ed), *Handbook of intelligence: Theories, measurements and applications.* New York, NY: Wiley.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (rev. ed., pp. 645-865). New York, NY: Academic Press.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443-451). New York, NY: Macmillan.

Horn, J. L. (1998). A basis of research on age differences in cognitive capabilities. In J. J. McArdle & R. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57-91). Chicago, IL: Riverside.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.), (pp. 41-68). New York, NY: The Guilford Press.

Horn, J. L., & Bramble, W. J. (1967). Second-order ability structure revealed in rights and wrongs scores. *Journal of Educational Psychology,* 58, 115-122.

Horn, J. L., & Cattell, R. B. (1966a). Age differences in fluid and crystallized intelligence. *Acta Psychologica,19,* 117-134.

Horn, J. L., & Cattell, R. B. (1966b). Age differences in primary mental abilities. *Journal of Gerontology,* 21, 210-220.

Horn, J. L., & Cattell, R. B. (1966c). Refinement & test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology,* 57, 253-70.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129. Abstract.

Horn, J. L., & Donaldson, G. (1980b). On the myth of intellectual decline in adulthood. *American Psychologists, 31,* 709-719.

Horn, J. L. & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 44-99). New York, NY: Cambridge University Press.

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

Horney, K. (1945). *Our inner conflicts: A constructive theory of neurosis.* New York, NY: Norton.

Hough, L. M. (1992). The big five personality variables – construct confusion: Description versus prediction. *Human Performance,* 5, 139-155.

Hough, L. M. (2001). I/Owes its advances to personality. In R. W. Roberts, & R. Hogan (Eds.), *Personality psychology in the workplace,* (pp. 19-45). Washington, DC: American Psychological Association.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581-595.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future – remembering the past. *Annual Review of Psychology, 51,* 631-664.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Hough, L. M., & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In K. R. Murphy (Ed.). *Individual differences and behaviour in organizations* (pp. 31-89). San Francisco: Jossey-Bass.

Hubley, A. M., Wu, A. D., & Zumbo, B. D. (2009, August). *Interpreting IRT parameters: putting psychological meat on the psychometric bone.* Presented at the annual meeting of the American Psychological Association, Toronto, ON.

Hülsheger, U. R., & Anderson, N. (2009). Applicant perspectives in selection: going beyond preference reactions. *International Journal of Selection and Assessment, 17,* 335-345.

Humphreys, L. G. (1986). Commentary on "the g factor in employment". *Journal of Vocational Behavior,* 29, 421-437.

Humphreys, L. M. (1973). Statistical definitions of test validity for minority groups. *Journal of Applied Psychology, 58,* 1-4.

Hunsley, J. & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446-455.

Hunt, E. (1995a, July-August). The role of intelligence in modern society: Are social changes dividing us into intellectual haves and have-nots? The question pushed aside in the 1970s is back, and the issues are far from simple. *American Scientist, 83,* 356-368.

Hunt, E. (1995b). *Will we be smart enough? A cognitive analysis of the coming workforce.* New York, NY: Sage.

Hunter, J. E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB).* Washington, DC: US Department of Labor, Employment Service.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational behavior, 29,* 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-95.

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

Ilgen, D R., & Pulakos, E. D. (1999). *The changing nature of performance: Implications for staffing, motivation, and development.* San Francisco: Jossey-Bass.

Ilies, R., Fulmer, I. S., Spitzmuller, M., & Johnson, M. D. (2009). Personality and citizenship behaviour: the mediating role of job satisfaction. *Journal of Applied Psychology, 94,* 945-959.

Ilies, R., Scott, B. A., & Judge, T. A. (2006). The interactive effects of personal traits and experienced states on intraindividual patterns of citizenship behavior. *Academy of Management Journal, 49,* 561-575.

Ip, E. H. (2000). Adjusting for information inflation due to local dependency in moderately large item clusters. *Psychometrika, 65,* 73-91.

Ip., E. H. (2001). Testing for local dependency in dichotomous and polytomous item response models. *Psychometrika, 66,* 109-132.

Isingrini, M. & Vazou, F. (1997). Relation between fluid intelligence and frontal lobe functioning in older adults. *Intelligence, Aging, and Human development, 45,* 99-109.

Jackson, D., Cooper-Thomas, H., van Gelderen, M., & Davis, J. (2010). Relationships among developmental competency measures and objective work outcomes in a New Zealand retail context. *Human Resource Development Quarterly, 21*, 169 – 186.

Jakobwitz, S., & Egan, V. (2006). The dark triad of normal personality traits. *Personality and Individual Differences, 40,* 331-339.

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1*, 131-163.

Jang, K. L., Livesley, J. W., Ando, J., Yamagata, S., Suzuki, A., Angleitner, A., Ostendorf, F., Riemann, R., & Spinath, F. (2006). Behavioral genetics of the higher-order factors of the big five. *Personality and Individual Differences, 41,* 261-272.

Jansen, M. G. H. (1997). Rasch's model for reading speed with manifest explanatory variables. *Psychometrika, 62,* 393-409.

Jattuso, M. L, & Sinar, E. F. (2003). Source effects in Internet-based screening procedures. *International Journal of Selection and Assessment, 11*, 137-140.

Jennrich, R. I., & Sampson, P. F. (1968). Application of stepwise regression to non-linear estimation. *Technometrics*, *10*, 63-72.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93-132). Heidelberg: Springer.

Jensen, A. R. (1998a). Jensen on "Jensenism". *Intelligence, 26,* 181-205.

Jensen, A. R. (1998b). *The g factor: The science of mental ability.* Westport, Connecticut: Paeger.

Jensen, A. R. (2002). Psychometric *g*: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds), *The general factor of intelligence: How general is it?* (pp. 39-53). New Jersey: Lawrence Erlbaum.

Jensen, A. R., & Rotter, J. B. (1947). The value of thirteen psychological tests in officer candidate screening. *Journal of Applied Psychology, 31,* 312-322.

Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 139-242). Norwood, NJ: Ablex.

Jewell, L. N., Siegall, M. (1990). *Contemporary industrial organisational psychology (*2nd ed.). New York, NY: West Publishing Co.

Johns, G. (2002). The psychology of lateness, absenteeism, and turnover. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 2: Organizational Psychology), (pp. 232-252). London: Sage.

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22,* 420-432.

Jones, J. W. (1991). Assessing privacy invesiveness of psychological test items: job relevant versus clinical measures of integrity. *Journal of business and Psychology, 5,* 531-535.

Jones, J. W., Joy, D. S., Werner, S. H., Orban, J. A. (1991). Criterion-related validity of a pre-employment integrity inventory: A large scale between groups comparison. *Perceptual and Motor Skills, 72,* 131-136.

Jones, J. W., Slora, K. B., Boye, M. W. (1990). Theft reduction through personnel selection: A control group design in the supermarket industry. *Journal of Business and Psychology, 5,* 275-279.

Jones, J. W., & Terris, W. (1983). Personality Correlates of Theft and Drug Abuse among Job Applicants. *Proceedings of the Third International Conference on the 16PF Test, 3*, 85-94.

Jones, W. H., Couch, L., & Scott, S. (1997). Trust and betrayal: The psychology of getting along and getting ahead. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 466-481). New York, NY: Academic Press.

Judge, T. A., Klinger, R., Simon, L. S., & Wen Fen Yang, I. (2008). The contributions of personality to organizational behaviour and psychology: findings, criticisms, and future research directions. *Social and Personality Psychology Compass, 2,* 1982-2000.

Jung, C. G. (1933). *Modern man in search of a soul.* New York, NY: Harcourt, Brace & World.

Juhasz, A. M., & Sonnenshein-Schneider, M. (1987). Adolescent sexuality: Values, morality, and decision making. *Adolescence, 22,* 579-590.

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review, 111,* 512-523.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence, 36,* 99-108.

Kanfer, R., Ackerman, P. L., Murtha, T., & Goff, M. (1995). Personality and intelligence in industrial and organisational psychology. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, (pp. 577-602). New York, NY: Plenum Press.

Kanfer, R., Wolf, M. B., Kantrowitz, T. M., Ackerman, P. L. (2010). Ability and trait complex predictors of academic and job performance: A person-situation approach. *Applied Psychology: An international Review, 59,* 40-69.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Sydney: Wadsworth.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Boston, MA: Allyn & Bacon.

Kaufman, A. S. (1993). Joint exploratory factor analysis of the Kaufman Assessment Battery for Children and the Kaufman Adolescent and Adult Intelligence Test for 11 and 12-year-olds. *Journal of Clinical Child Psychology, 22,* 355-364.

Kaufman, A. S., Kaufman, J. C., & Mclean, J. E. (1995). Factor structure of the Kaufman Adolescent and Adult Intelligence Test (KAIT) for whites, African Americans, and Hispanics. *Educational and Psychological Measurement, 55,* 365-376.

Kaufman, J. D., & Borman, W. C. (2004). Citizenship performance in organisations. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 412-426). New Jersey: Wiley.

Kelloway, E. K., Francis, L., Prosser, M., & Cameron, J. E. (2010). Counterproductive work behavior as protest. *Human Resource Management Review, 20,* 18-25.

Kelloway, E. K., Loughin, C., Barling, J., & Nault, A. (2002). Self-reported counterproductive behaviors and organizational citizenship behaviors: Separate but related constructs. *International Journal of Selection and Assessment, 10,* 143-151.

Kelly, G. A. (1963). *A theory of personality: The psychology of personal constructs.* New York: Norton.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy. *American Psychologist, 43,* 23-261.

Killeen, P. R. (2001). Writing and overwriting short-term memory. *Psychonomic Bulletin & Review*, *8*, 18-43.

Kingsbury, G. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education, 4,* 241-26

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurements, 56,* 746-759.

Kisamore, J. L., Stone, T. H., & Jawahar, I. M. (2007). Academic integrity: the relationship between individual and situational factors on misconduct contemplations. *Journal of Business Ethics, 75,* 381-394.

Kline, P. (1995). A critical review of the measurement of personality and intelligence. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, (pp. 505-524). New York: Plenum Press.

Kline, P. (1992). *Handbook of psychological testing.* London: Routledge.

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

Kline, P. (2000). *Handbook of psychological testing (2nd ed.).* London: Routledge.

Kobbs, S. W., & Arvey, R. D. (1993). Distinguishing deviant and non-deviant nurses using the Personnel Reaction Blank. *Journal of Business and Psychology, 8,* 255-264.

Kong, X. J., Wise, S. L., & Bhola, D. S. (2007). Setting the response time threshold parameter to differentiate solution behaviour from rapid-guessing behavior. *Educational and Psychological Measurement, 67,* 606-619.

Korchkin, S. (1987). Personality correlates of a measure of honesty. *Journal of Business and Psychology, 1,* 236-247.

Kornhauser, A. W. (1922). The psychology of vocational selection. *Psychological Bulletin, 19,* 192-229.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking big personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological Bulletin, 136,* 768-821.

Kouzes, J. M., & Posner, B. Z. (1993). *Credibility: How leaders gain and lose it, why people demand it.* San Francisco: Jossey-Bass.

Kramer, R. M., & Cook, K. S. (2004). *Trust and distrust in organizations: Dilemmas and approaches.* New York, NY: Sage.

Kroeck, K. G., & Brown, K. W. (2004). Work applications of the big five model of personality. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 109-130). Mahwah, NJ: Wiley.

Kroeck, K. G., & Fraser, S. L. (1991). Comparison of polygraph and honesty inventory results. *Forensic Reports, 4,* 225-233.

Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: the use of idiosyncratic item responses. International Journal of Selection and Assessment, 15, 220-231.

Kurtz, A. K. (1948). A research test of the Rorschach test. *Personnel Psychology, 1,* 41-51.

Kyllonen, P. (1997). Smart testing. In R. Dillon (Ed.), *Handbook on testing* (pp. 347-368). Westport, CT: Greenwood.

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

Ladd, D., & Henry, R. A. (2000). Helping coworkers and helping the organization: The role of support perceptions, exchange ideology, and conscientiousness. *Journal of Applied Social Psychology, 30,* 2028-2049.

Lakatos, I. (1968). Criticism and the methodology of scientific research programmes. *Proceedings of the Aristotelian Society, 69,* 149-186.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9,* 202-220.

Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41,* 1183-1192.

Landy, F. J., Shankster, L. J., Kohler, S. S. (1994). Personnel selection and placement. *Annual Review of Psychology, 45,* 261-96.

Laland, K. N. (2003). Gene-culture co-evolution. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 1-7). London: Nature Publishing Group.

Lautenschlager, G., & Flaherty, V. (1990). Computer administration of questions: More desirable or more socially desirable? *Journal of Applied Psychology, 75*, 310-314.

Lavoisier, A. L. (1789). Elements of chemistry (translated by R. Kerr). In R. M. Hutchins (Ed.), *Lavoisier, Fourier, Faraday* (vol.45) – *Great books of the Western world* (pp. 1-169). London: Encyclopaedia Britannica, Inc.

Lazarus, R. S. (1991). Psychological stress in the workplace. In P. L. Perrewe (Ed.), *Journal of Social Behavior and Personality: Vol. 6. Handbook on job stress* (special issue) (pp. 1–13). San Rafael, CA: Select Press.

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology, 98*, 668-682.

Leong, F. T. L., & Austin, J. T. (1996). *The psychology research handbook: A guide for graduate students and research assistants*. London: Sage.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86,* 326-336.

Leonard, J. A. (1996). *Response latency as an alternative measure of performance on honesty tests.* Unpublished doctoral dissertation, University of South Florida. Abstract.

Levashina, J., Morgeson, F. P., & Campion, M. A. (2009). They don't do it often, but they do it well: exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Asessment, 17*, 271-281.

Levine, E. L. (2010). Emotion and power (as social influence): their impact on organizational citizenship and counterproductive individual and organizational behavior. *Human Resource Management Review, 20,* 4-17.

Levy-Leboyer, C. (1989). A case study on examining the accuracy of predictors. In M. Smith & I. T. Robertson (Eds.). *Advances in selection and assessment* (pp. 129-132). Chichester, United Kingdom: Wiley.

Lewicki, R. J., & Wiethoff, C. (2000). Trust, trust development, and trust repair. In M. Deutsch & P. Coleman (Eds.), *The handbook of conflict resolution: Theory and practice* (pp. 86-108). San Francisco: Jossey-Bass.

Lewis, J. D. (1999). *Trusted partners: How companies build mutual trust and win together*. New York, NY: The Free Press.

Licht, M. H. (2003). Multiple regression and correlation. In L. G. Grimm & P. R. Yarnold, *Reading and understanding multivariate statistics* (pp. 19-65). Washington, DC: American Psychological Association.

Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment, 10,* 245-257.

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: ordinal/rating scale item responses. *Educational and Psychological Measurement, 70,* 5-21. Retrieved from http://epm.sagepub.com/cgi/content/abstract/70/1/5

LoBello, S. G., & Sims, B. N. (1993). Fakability of a commercially produced pre-employment integrity test. *Journal of Business and Psychology, 8,* 265-273.

Loehlin, J. C., Lindzey, G., & Spuhler, J. M. (1975). *Race differences in intelligence.* San Francisco: Freeman.

Logan, T. G., Koettel, R. C., & Moore, R. W. (1986). Personality correlated of a test of honesty. *Psychological Reports, 59,* 1015-1018.

Lohman, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research, 59,* 333-373.

Lohman, D. F., & Rocklin, T. (1995). Current and recurring issues in the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, pp. (447-474). New York, NY: Plenum Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Welsley Publishing Company.

Lounsbury, J., Borrow, W., & Jensen, J. (1989). Attitudes toward employment testing: Scale development, correlates, and "known-group" validation. *Professional Psychology: Research and Practice, 20*, 340-349.

Lovallo, W. R. (1997). *Stress & Health: Biological and psychological interactions*. Thousand Oaks, CA: Sage.

Lovell, R., & Norris, M. (1990). *One in four: Offending from age ten to twenty-four in a cohort of New Zealand males* (Study of social adjustment: Research Report No 8). Wellington, New Zealand: Department of Social Welfare.

Lowman, R. L. (1996). Work dysfunctions and mental disorders. In K. R. Murphy (Ed.), *Individual differences and behaviour in organizations* (pp. 371-416). San Francisco: Jossey-Bass.

Lubart, T. I. (2003). In search of creative intelligence. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: international perspectives*, (pp. 279-293). Washington, D.C: American Psychological Association.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York, NY: Oxford University Press.

Manly, B. F. J. (2005). *Multivariate statistical methods: A primer.* Boca Raton, FL: Chapman & Hall/CRC.

March, J. (1998). *Dictionary of classical mythology*. London: Cassell.

Marcus, B., Hoft, S., & Riediger, M. (2006). Integrity tests and the five-factor model of personality: A review and empirical test of two alternative positions. *International Journal of Selection and Assessment, 14,* 113-130.

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: big five, or one in addition? *Personnel Psychology, 60,* 1-34.

Marcus, B., Wagner, U., Poole, A., Powell, D. M., & Carswell, J. (2009). The relationship of GMA to counterproductive work behavior revisited. *European Journal of Personality, 23,* 489-507.

Marks, G., & Miller, N. (1987). Ten years of research on the false consensus effect: An empirical and theoretical review. *Psychological Bulletin, 102,* 72-90.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radix and hierarchical models of intelligence. *Intelligence,* 7, 107-127.

Martin, S. L., & Terris, W. (1991). Predicting infrequent behavior: Clarifying the impact on false-positive rates. *Journal of Applied Psychology, 76,* 484-487.

Martin, S. L., & Terris, W. (1994). Re-emphasizing Martin and Terris's (1991) "Predicting infrequent behavior: Clarifying the impact of false-positive rates." *Journal of Applied Psychology, 79,* 302-305.

Martinko, M. J., Douglas, S. C., & Harvey, P. (2006). Understanding and managing workplace aggression. *Organizational Dynamics, 35,* 117-130.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52,* 397-422.

Maslow, A. H. (1970). *Motivation and personality* (2nd ed). New York, NY: Harper & Row.

Matlock-Hetzel, S. (1997). *Basic Concepts in Item and Test Analysis.* Retrieved from http://ericae.net/ft/tamu/Espy.htm.

Matthews, G., & Deary, I. J. (1998). *Personality traits*. London: Cambridge University Press.

Maxwell, S. E., Delaney, H. D. (2004). *Designing experiments and analysing data: A model comparison perspective (2nd ed.).* London: Lawrence Erlbaum.

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality*, 60, 329-361.

McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamic hypotheses of Gf-Gc theory using longitudinal life-span data. *Learning and Individual Differences, 12,* 53-79.

McClelland, D. C. (1975). *Power: The inner experience.* New York, NY: Irvington Press.

McClelland, D. C., & Franz, C. E. (1993). Motivational and other sources of work accomplishments at mid-life: A longitudinal study. *Journal of Personality, 60,* 679-707.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and questionnaires. *Journal of Personality and Social Psychology, 49,* 710-721.

McCrae, R. R., & Costa, P. T. (1986). Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist, 41,* 1001-1002.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81-90.

McCrae, R. R., & Costa, P. T. (1989). The structure of interpersonal traits: Wiggin's circumplex and the five-factor model. *Journal of Personality and Social Psychology, 56,* 586-595.

McCrae, R. R., & Costa, P. T. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. Johnson, & S. Briggs (Eds), *Handbook of Personality Psychology* (pp. 826-849)*.* San Diego: Academic Press.

McCrae, R. R., Costa, P. T., de Lima, M. P., Simoes, A., Ostendorf, F., Angleitner, A., et al. (1999). Age differences in personality across the adult life span: parallels in five cultures. *Developmental Psychology,* 35, 466-477.

McCrae, R. R., Costa, P. T., & Del Pilar, G. H. (1998). Cross-cultural assessment of the five-factor model: The revised NEO personality inventory. *Journal of Cross-Cultural Psychology,* 29(Special issue 1: Personality and its measurement in cross-cultural perspective), 171-188.

McCrae, R. R., Costa, P. T., Ostendorf, F., A., Hrebickova, M., Avia, M., Sanz, J., et al. (2000). Nature over nurture: Temperament, personality, and lifespan development. *Journal of Personality and Social Psychology,* 78, 173-186.

McCrae, R. R., Costa, P. T., & Terracciano, A. (2002). Personality trait development from age 12-18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of Personality and Social Psychology,* 83, 1456-1468.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality,* 60, 175-215.

McCrae, R. R., Yamagata, S., Jang, K. L., Riemann, R., Ando, J., Ono, Y., Angleitner, A., & Spinath, F. M. (2008). Substance and artefact in the higher-order factors of the big five. *Personality Processes and Individual Differences, 95,* 442-455.

McCrae, R. R., Yik, M. S. M., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: bilingual, acculturation, and peer rating studies of Chinese Undergraduates. *Journal of Personality and Social Psychology, 74,* 1041-1055.

McDaniel, M. A. (2009). Gerrymandering in personnel selection: a review of practice. *Human Resource Management Review, 19,* 263-270.

McDaniel, M. A., & Jones, J. W. (1988). Predicting employee theft: A quantitative review of the validity of a standardised measure of dishonesty. *Journal of Business and Psychology, 2,* 327-343.

McDaniel, M. A., & Timm, T. (1990). *Lying Takes Time: Predicting Deception in Biodata using Response Latency.* Paper presented at the 98th Annual conference of the American Psychological Association, Boston, MA.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8,* 290-302.

McFarland, R. A. (1930). An experimental study of the relationship between speed and mental ability. *Journal of General Psychology, 3,* 67-97.

McGhee, R., & Lieberman, L. (1994). Gf-Gc theory of human cognition: Differentiation of short-term auditory and visual memory factors. *Psychology in the Schools, 31,* 297-304.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf-Gc* framework of human cognitive and knowledge abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 151-180). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.), (pp. 136-181). New York, NY: The Guilford Press.

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

McGrew, K. S., & Flanagan, D. P. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 189-207.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136,* 450-470.

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.

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

McNemar, Q. (1964). Lost: our intelligence? Why? *American Psychologist,* 19, 871-882.

Mehrabian, A. (2000). Beyond IQ: Broad-based measurement of individual success potential or "emotional intelligence." *Genetic, Social, and General Psychological Monographs, 126,* 133-239.

Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology, 55,* 675-680.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30,* 955-966.

Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's frames of mind and Sternberg's beyond IQ in the context of theory and research on the structure of human abilities. *Psychological Inquiry, 3,* 365-384.

Meyer, J. P. (2008, March). *A mixture Rasch model with item response time components*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Mikulay, S. M., & Goffin, R. D. (1998). Measuring and predicting counterproductivity in the laboratory using integrity and personality testing. *Educational and Psychological Measurement, 58,* 768-791.

Miller, R. B. (1966). Human factors in systems. In B. V. H. Glimer. *Industrial psychology,* (pp. 218-236). New York, NY: McGraw-Hill.

Millman, J. (1969). *How to take tests.* New York, NY: McGraw-Hill.

Miner, J. B. (1983). The unpaved road from theory: Over the mountains to application. In R. H. Kilmann, K. W. Thomas, D. P., Slevin, R. Nath, & S. L. Jerrel (Eds.), *Producing useful knowledge for organizations* (pp. 37-68). New York, NY: Praeger.

Mischel, W. (1968). *Personality and assessment.* New York, NY: Wiley.

Mischel, W. (1969). Continuity and change in personality. *American Psychologist, 24,* 1012-1018.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80,* 252-283.

Mischel, W. (1981). *Introduction to personality (3rd ed.).* New York, NY: Holt, Rinehart and Winston.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55,* 1-22.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59,* 439-483.

Moore, R. W. (1990). Instructional effects on the Phase II Profile honesty test. *Psychological Reports, 67,* 291-294.

Moretti, D. M., & Terris, W. (1983). *The Personnel Selection Inventory: A predictive validation study.* Park Ridge, IL: London House Press.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.

Moscoso, S., & Iglesias, M. (2009). Job experience and big five personality dimensions. *International Journal of Selection and Assessment, 17,* 239-242.

Moscoso, S., & Salgado, J. F. (2004). "Dark side" personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment, 12,* 356-362.

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.

Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resources management. In K. M. Rowland & G. Ferris (Eds.), *Research in personnel and human resources management* (pp. 153-200). Greenwich, CT: JAI Press.

Mount, M. K., Barrick, M. R, & Stewart, G. L. (1998). Five-factor model of personality and job performance in jobs involving interpersonal interactions. *Human Performance, 11,* 145-165.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher-order dimensions of the big five personality traits and the big six vocational interest types. *Personnel Psychology, 58*, 447–478.

Moutafi, J., Furnham, A., & Crump, J. (2003). Demographic and Personality Predictors of Intelligence: A study using the Neo Personality Inventory and the Myers-Briggs Type Indicator. *European Journal of Personality,* 17, 79-94.

Moutafi, J., Furnham, A., & Paltiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences, 37,*1013-1022.

Moutafi, J., Furnham, A., & Paltiel, L. (2005). Can personality factors predict intelligence? *Personality and Individual Differences, 38,*1021-1033.

Murphy, K. R. (1986). When your top choice turns you down: effect of rejected job offers on the utility of selection tests. *Psychological Bulletin, 99,* 133-138.

Murphy, K. R. (1993). *Honesty in the workplace.* Pacific Grove, California: Brooks/Cole Publishing Company.

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than g. In K. R. Murphy (Ed.), *Individual differences and behaviour in organizations* (pp. 3-31). San Francisco: Jossey-Bass.

Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (5th ed.). Mahwah, NJ: Prentice Hall International, Inc.

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18,* 343-357.

Murphy, K. R., & Lee, S. L. (1994a). Does conscientiousness explain the relationship between integrity and job performance? *International Journal of Selection and Assessment, 2,* 226-233.

Murphy, K. R., & Lee, S. L. (1994b). Personality variables related to integrity scores: The role of conscientiousness. *Journal of Business and Psychology, 8,* 413-424.

Murphy, K. R., Myors, B., & Wolach, A. (2009). *Statistical power analysis: A simple and general model for traditional and modern hypothesis testing* (3rd ed.). NY, New York: Routledge.

Necka, E. (2003). Intelligence and temperament. In R. J. Sternberg, J. Lautrey, &T. I. Lubart (Eds), *Models of intelligence: international perspectives*, (pp. 293-309). Washington, D.C: American Psychological Association.

Neisser, U., et al., (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77-101.

Nesselroade, J. R., & Cattell, R. B. (Eds.) (1988). *Handbook of multivariate experimental psychology.* New York, NY: Plenum.

Neubauer, A. C. (1990). Speed of information processing in the Hick paradigm and response latencies in a psychometric intelligence test. *Personality and Individual Differences, 11,* 147-152.

Neubauer, A. C. (1997). The mental speed approach to the assessment of intelligence. In J. Kingma, & W. Tomic (Eds.), *Advances in cognition and educational practice: Reflections on the concept of intelligence* (pp. 149-174). Greenwich, Connecticut: JAI Press.

Newbold, G. (2000). *Crime in New Zealand.* Palmerston North, New Zealand: Dunmore Press Limited.

New Zealand Press Association (2005, January 27). Law change will give bad workers jobs for life. *The Dominion Post,* p. A4.

Noll, J. G., & Horn, J. L. (1998). Age differences in processes of fluid and crystallized intelligence. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice,* (pp. 263-281). Mahwah, NJ: Lawrence Erlbaum.

Norman, W. T. (1963). Toward and adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.

Norman, S. B. (2003). The effect of training managers to gain employee trust on employee work-related attitudes and emotional well-being. *Dissertation Abstracts International:* UMI No. 3090648.

Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*, 1-18.

Nunnally, J.C. (1972). *Educational measurement and evaluation* (2nd ed.). New York, NY: McGraw-Hill.

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

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Nyborg, H. (Ed.) (2003). *The scientific study of general intelligence: A tribute to Arthur R. Jensen.* Amsterdam: Pergamon Press.

Nye, C. D., Do, B., Drasgow, F., & Fine, S. (2008). Two-step testing in employee selection: is score inflation a problem? *International Journal of Selection and Assessment, 16*, 112-120.

O'Bannon, R. M., Goldinger, L. A., & Appleby, G. S. (1989). *Honesty and integrity testing.* Atlanta, GA: Applied Information Resources.

Ones, D. S. (1993). The construct validity of integrity tests. *Dissertation Abstracts International:* UMI No. 9404524. Abstract

Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organization Psychology, 75,* 255-276.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60,* 995-1027.

Ones, D. S, Schmidt, F. L., & Viswesvaran, C. (1993a, May). *Integrity and Ability: Implications for Incremental Validity and Adverse Impact.* Paper presented at the Eighth annual conference of the Society of Industrial and Organizational Psychology, San Francisco.

Ones, D. S, Schmidt, F. L., & Viswesvaran, C. (1993b, May). *Nomological Net for Measures of Integrity and Conscientiousness.* Paper presented at the Eighth annual conference of the Society of Industrial and Organizational Psychology, San Francisco.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609-626.

Ones, D. S., & Viswesvaran, C. (1998a). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83,* 35-42.

Ones, D. S., & Viswesvaran, C. (1998b). Integrity testing in organizations. In R. W. Griffin, A. O'Leary-Kelly, & J. M. Collins (Eds.), *Dysfunctional behavior in organizations: Violent and deviant behavior,* (pp. 243-276). New York, NY: JAI Press.

Ones, D. S., & Viswesvaran, C. (1998c). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11,* 245-269.

Ones, D. S., & Viswesvaran, C. (2001a). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9,* 31-39.

Ones, D. S., & Viswesvaran, C. (2001b). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace,* (pp. 63-92). Washington, DC: American Psychological Association.

Ones, D. S., & Viswesvaran, C. (2003). The big-5 personality and counterproductive behaviours. In A. Sagie, S. Stashevsky, & M. Koslowsky (Eds.), *Misbehaviour and dysfuncational attitudes in organizations*, (pp. 211-249). Hampshire, UK: Palgrave/MacMillan.

Ones, D. S., & Viswesvaran, C. (2007). Labor market influences on personality scale scores among job applicants: four field studies in personnel selection settings. *Zeitschrift fur Personalpsychologie, 6*, 71-84.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005a). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection,* (pp. 143-173). Malden, MA: Blackwell Publishing.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005b). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence,* (pp. 431-468). Thousand Oaks, CA: Sage.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005c). Personality at work: Raising awareness and correcting misconceptions. *Human Performance, 18,* 389-404.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660-679.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1992, May). *An Examination of Validity Generalization and Moderators of Integrity Test Validity.* Presented at Annual Conference of Society for Industrial and Organizational Psychology, Montreal.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679-703.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1995). Integrity tests: Overlooked facts, resolved issues, and remaining questions. *American Psychologist, 50,* 456-457.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Psychology, 17,* S19-S38.

Organ, D. W. (1994). Personality and organizational citizenship behavior. *Journal of Management, 20,* 465-478.

Organ, D. W. (1997). Organizational Citizenship Behavior: It's construct clean-up time. *Human Performance, 10,* 85-97.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48,* 775-802.

Ortner, T. M. (2008). Effects of changed item order: a cautionary note to practitioners on jumping to computerized adaptive testing for personality assessment. *International Journal of Selection and Assessment, 16*, 249-257.

Osborn, R. N., & Marion, R. (2009). Contextual leadership, transformational leadership and the performance of international innovation seeking alliances. *The Leadership Quarterly, 20,* 191-206.

Osborne, J. & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8*. Retrieved from http://PAREonline.net/getvn.asp?v=8&n=2

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

Packman, T., Brown, G. S., Englert, P., Sisarich, H., & Bauer, F. (2005). Differences in personality traits across ethnic groups within New Zealand and across an international sample. *New Zealand Journal of Psychology, 34,* 77-85.

Pallier, G., Roberts, R. D., & Stankov, L. (2000). Biological versus psychometric intelligence: Halstead's (1947) distinction revisited. *Archives of Clinical Neuropsychology, 15,* 205-226.

Parshall, C. G., Mittelholtz, D., & Miller, T. R. (1994, April). *Response time: An investigation into determinants of item-level timing.* Paper presented at the meeting of the National Council of Measurement in Education, New Orleans, LA.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36,* 556-563.

Payne, R. (1999). Stress at work: A conceptual framework. In J. Firth-Cozens & R. Payne (Eds.), *Stress in health professionals: Psychological and organisational causes and interventions* (pp. 3–16). Chichester, UK: Wiley.

Pearce, R. (2009). Article review: personality assessment in organizational settings. *Griffith University Undergraduate Student Psychology Journal, 1,* 5-10.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* London: Lawrence Erlbaum.

Penney, L. M., & Borman, W. C. (2005). The prediction of contextual performance. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection,* (pp. 377-396). Malden, MA: Blackwell Publishing.

Penney, L. M., & Spector, P. E. (2005). Job stress, incivility, and counterproductive work behavior (CWB): the moderating role of negative affectivity. *Journal of Organizational Behavior, 26,* 777-796.

Penney, L. M., Spector, P. E., & Fox, S. (2003). Stress, personality and counterproductive work behaviour. In A Sagie, S. Stashevsky, & M. Koslowsky (Eds.), *Misbehavior and dysfunctional attitudes in organizations,* (pp. 194-210). Hampshire, UK: Palgrave/MacMillan.

Petrill, S. A., & Deary, I. J. (2001). Inspection time and intelligence celebrating 25 years of research. *Intelligence, 29,* 441-442.

Pettit, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers, 34*, 50-54.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks, CA: Sage.

Piaget, J. (1973). *The child and reality.* New York, NY: Grossman.

Piedmont, R.. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity studies: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582-593.

Plato. (1999). Republic (translated by P. Shorey). In E. Hamilton & H. Cairns (Eds.), *The collected dialogues of Plato: Including the letters*. Mahwah, NJ: Princeton University Press. (Original work published 380 BCE).

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual and organizational level consequences of organizational citizenship behaviours: a meta-analysis. *Journal of Applied Psychology, 94,* 122-141.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26,* 513-563.

Postlethwaite, B., Robbins, S., Rickerson, J., & McKinniss, T. (2009). The moderation of consicnetiousness by cognitive ability when predicting workplace safety behavior. *Personality and Individual Differences, 47,* 711-716.

Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: corrected estimates and realistic expectations. *International Journal of Selection and Assessment, 13,* 304-315.

Potosky, D., Bobko, P., & Roth, P. L. (2008). Some comments on pareto thinking, test validity, and adverse impact: when ‘and’ is optimal and ‘or’ is a trade-off. *International Journal of Selection and Assessment, 16,* 201-205.

Powers, D. E., & Kaufman, J. C. (2004). Do standardized tests penalize deep-thinking, creative, or conscientious students? Some personality correlates of Graduate Record Examinations test scores. *Intelligence, 32,* 145-153.

Psychometrics Ltd. (2002). *The 15FQ+ technical manual.* London: Psychometrics Ltd.

Psytech International Ltd. (2009). *Graduate reasoning test (GRT1)/General reasoning test (GRT2)/Critical reasoning test battery (CRTB1).* London: Psytech International Ltd.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241-258.

Rabbitt, P., & Abson, V. (1991). Do older people know how good they are? *British Journal of Psychology, 82,* 233-238.

Rafilson, F. M., & Frost, A. G. (1989). Overt integrity tests versus personality-based measures of delinquency: An empirical comparison. *Journal of Business and Psychology, 3,* 269-277.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: The University of Chicago Press.

Rawlings, D., & Carnie, D. (1989). The interaction of EPQ extraversion with WAIS subtest performance under timed and untimed conditions. *Personality and Individual Differences, 10,* 453-458.

Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology, 30,* 59-64.

Rayburn, J. M., & Rayburn, L. G. (1996). Relationship between Machiavellianism and Type A personality and ethical-orientation. *Journal of Business ethics, 15,* 1209-1219.

Read, S.J., Monroe, B.M., Brownstein, A.L., Yang, Y., Chopra, G., & Miller, L.C. (2010) A neural network model of the structure and dynamics of human personality. *Psychological Review, 117*, 61-92.

Reader, S. M. (17 July, 2004). Don't call me clever. *New Scientist,* 35-37.

Reber, A. S., & Reber, E. (2001). *The Penguin dictionary of psychology* (3rd ed). Auckland, NZ: Penguin Books.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology, 44,* 321-332.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1,* 86-89.

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.

Reeve, C. L. (2007). Functional impulsivity and speeded ability test performance. *International Journal of Selection and Assessment, 15*, 56-62.

Reeve, C. L., Meyer, R. D., & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence, 34,* 387-402.

Reina, D. S., & Reina, M. L. (1999). *Trust and betrayal in the workplace: Building effective relationships in your organization.* San Francisco: Berrett-Koehler Publishers, Inc.

Rest, J. R. (1984). The major components of morality. In W. M. Kurtines & J. L. Gerwitz (Eds.), *Morality, moral behavior, and moral development* (pp. 24-38). New York, NY: Wiley.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the GLB: comments on Sijtsma. *Psychometrika, 74,* 145-154.

Richardson, K (1991). *Understanding intelligence.* Buckingham, UK: Open University Press.

Richardson, K. (1998). *The origins of human potential: Evolution, development, and psychology.* New York, NY: Routledge.

Richardson, K., & Spears, D. (1972). *Race, culture and intelligence.* Harmondsworth: Penguin.

Riggio, R. E. (2000). *Introduction to industrial/organizational psychology* (3rd ed.). Mahwah, NJ: Prentice Hall.

Rindermann, H., & Neubauer, A. C. (2001). The influence of personality on three aspects of cognitive performance: processing speed, intelligence and school performance. *Personality and Individual Differences, 30,* 829-842.

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB) little more than acculturated learning (Gc)!? *Learning and Individual Differences, 12,* 81-103.

Roberts, R. D., Markham, P. M., Matthews, G., & Zeidner, M. (2005). Assessing intelligence: Past, present, and future. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence,* (pp. 333-360). Thousand Oaks, CA: Sage.

Roberts, B. W., Harms, P. D., Caspi, A., & Moffitt, T. E. (2007). Predicting the counterproductive employee in a child-to-adult prospective study. *Journal of Applied Psychology, 92,* 1427-1436.

Roberts, B. W., & Hogan R. (2001). *Personality psychology in the workplace*. Washington, DC: American Psychological Association.

Robinson, M. D., Wilkowski, B. M., & Meier, B. P. (2008). Approach, avoidance, and self-regulatory conflict: An individual differences perspective. *Journal of Experimental Social Psychology, 44*, 65-79.

Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy.* Boston, MA: Houghton Mifflin.

Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden & R. K. Hambleton (eds.), *Handbook of modern item response theory* (pp. 187-208). New York, NY: Springer.

Rosnow, R., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276-1284.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13,* 279-301.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54,* 297-330.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: a policy-capturing approach. *Journal of Applied Psychology, 87,* 66-80.

Rouder, J. N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika, 70,* 377-381.

Rupp, D. E., & Spencer, S. (2006). When customers lash out: the effects of customer interactional injustice on emotional labour and the mediating role of discrete emotions. *Journal of Applied Psychology, 91,* 971-978.

Rushton, J. P., & Irwing, P. (2008). A General Factor of Personality (GFP) from two meta-analyses of the Big Five: Digman (1997) and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences, 45,* 679-683.

Rushton, J. P., & Irwing, P. (2009). A general factor of personality in the Comrey Personality Scales, the Minnesota Multiphasic Personality Inventory-2, and the Multicultural Personality Questionnaire. *Personality and Individual Differences, 46*, 437-442.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52,* 359-392.

Ryan, A. M., & Sackett, P. R. (1987). Pre-employment honesty testing: Fakability, reactions of test takers, and company image. *Journal of Business and Psychology, 1,* 248-256.

Ryckman, R. M. (1993). *Theories of personality* (5th ed.). Los Angeles, CA: Brooks/Cole Publishing Company.

Rynes, S. L., & Connerly, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology, 7,* 261-275.

Sackett, P. R. (1994). Integrity testing for personnel selection. *Current Directions in Psychological Science, 3,* 73-76.

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

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

Sackett, P. R., Burris, L. R., & Callahan, C. (1989). integrity testing for personnel selection: An update. *Personnel Psychology, 42,* 491-529.

Sackett, P. R., De Corte, W., & Lievens, F. (2008). Pareto-optimal predictor composite formation: a complementary approach to alleviating the selection quality/adverse impact dilemma. *International Journal of Selection and Assessment, 16,* 206-209.

Sackett, P. R., & DeVore, C. J. (2001). Counterproductive behaviours at work. In N. Anderson, D. Ones, H. Sinangil, & C. Viswesvaran (Eds.), *International handbook of work psychology* (Vol. 1, pp. 145-164). London: Sage.

Sackett, P. R., Ellingson, J. E. (1997). The effects of forming mutli-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

Sackett, P. R., & Lievens, F. (2008). Personnel Selection. *Annual Review of Psychology, 59,* 419-450.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education – Prospects in a post-affirmative action world. *American Psychologist, 56,* 302-318.

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49,* 787-829.

Sackett, P. R., & Wilk, S. I. (1994). Within-group Norming and other forms of score adjustment in pre-employment testing. *American Psychologist, 49,* 929-954.

Saklofske, D. H., & Zeidner, M. (1995). *International handbook of personality and intelligence.* New York, NY: Plenum Press.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82,* 30-43.

Salgado, J. F. (1999). Personnel selection methods. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 1-54). London: Wiley.

Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment, 11*, 194-205.

Salkind, N. J. (2006). *Tests and measures for people who (think they) hate tests and measurement.* Kansas: Sage.

Salthouse, T. A. (1987). The role of representations in age differences in analogical reasoning. *Psychology and Aging, 2,* 357-362.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403-428. Abstract.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13,* 445-461.

Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1990). Age, self-assessed health status, and cognition. *Journal of Gerontology, 45,* 156-160.

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45,* 73-103.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study.* Cambridge, UK: Cambridge University Press.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: why there cannot be a debate. *Human Performance, 15*, 187-210.

Schmidt, F. L., Berner, J. G., & Hunter, J. E. (1973). Racial differences in validity of employment tests: Reality or illusion? *Journal of Applied Psychology, 58,* 5-9.

Schmidt, F. L., & Hunter, J. E. (1974). Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. *American Psychologist, 29,* 1-8.

Schmidt, F. L., & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology, 68,* 407-415.

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

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

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). The impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64,* 609-626.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). The impact of job experience and ability of job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432-439.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology, 66,* 166-185.

Schmidt, F. L., Mack, M. J., & Hunter, J. E. (1984). Selection utility in the occupation of U.S. Park Ranger for three modes of test use. *Journal of Applied Psychology, 69,* 490-497.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627-670.

Schmidt, F. L., Shaffer, J. A., & Oh, I. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61,* 827-868.

Schmidt, F. L., Viswesvaran, C., & Ones, D. S. (1997). Validity of integrity tests for predicting drug and alcohol abuse: A meta-analysis. In W. J. Bukoski (Ed.), *Meta-analysis of drug abuse prevention programs* (pp. 69-95). Rockville, MD: NIDA press.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82,* 719-730.

Schmitt, T. A., Sass, D. A., Sullivan, J. R., & Walker, C. M. (2010). A Monte Carlo simultation investigating the validity and reliability of ability estimation in Item Response Theory with speeded computer adaptive tests. *International Journal of Testing, 10,* 230-261.

Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits, or how to sink science in five dimensions or less. *Journal of Organizational Behavior, 17,* 639-655.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement, 34,* 213-232.

Schnipke, D. L., & Scrams, D. J. (1998, September). *Exploring issues of examinee behaviour: Insights gained from response time analysis.* Paper presented at the ETS colloquium on Computer based testing: building the foundation for future assessment, Philadelphia.

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., & Barta, P. (2000). Elucidating the contributions of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. *Journal of the International Neuropsychological Society, 6,* 52-61.

Scrams, D. J., & Schnipke, D. L. (1997, March). *Making use of response times in standardized tests: Are accuracy and speed measuring the same thing?* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Sechrest, L. (1963). Incremental validity: a recommendation. *Educational and Psychological Measurement, 23,* 153-158.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research,* 4, 27-41.

Shaw, R. B. (1997). *Trust in the balance: Building successful organizations on results, integrity, and concern*. San Francisco: Jossey-Bass.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2003). *Research methods in psychology.* New York, NY: McGraw-Hill.

Sheldon, W. H., & Stevens, S. S. (1942). *The varieties of temperament: A psychology of constitutional differences.* New York, NY: Harper & Brothers Publishers.

Shi, J., Lin, H., Wang, L., & Wang, M. (2009). Linking the big five personality constructs to organizational justice. *Social Behavior and Personality, 37,* 209-222.

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). Chicago, IL: University of Chicago Press.

Shu-chen, L., Jordanova, M., & Lindenberger, U. (1998). From good senses to good sense: A link between tactile information processing and intelligence. *Intelligence, 26,* 99-122.

Siem, F. M. (1998). Metatraits and self-schemata: same or different? *Journal of Personality, 66,* 783-803.

Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107-120.

Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Pscyhometrika, 74,* 169-173.

Simon, H. A. (1997). *Models of bounded rationality: Vol.3. Empirically grounded economic reason.* Cambridge, MA: MIT Press.

Simon, R. J., & Mahan, L. (1971). Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law & Society Review, 5,* 319-330.

Simonini, S. L. (1998). Investigating the relationship between integrity and job turnover. *Dissertation Abstracts International:* UMI No. 1390747.

Skinner, B. F. (1965). *Science and human behavior.* New York, NY: Free Press.

Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment, 15,* 467-477.

Smith, D. B., & Ellington, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87,* 211-219.

Smith, E. A., Borgerhoff, Mulder, M., & Hill, K. (2001). 'Controversies in the evolutionary social sciences: A guide for the perplexed'. *Trends in Ecology and Evolution, 16,* 128-134.

Smith, G. T. (2005a). On construct validity: Issues of method and measurement. *Psychological Assessment, 17,* 396-408.

Smith, G. T. (2005b). On the complexity of quantifying construct validity. *Psychological Assessment, 17,* 413-414.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*, 49-76.

Snow, R. E. (1986). Individual differences and the design of educational programs. *American Psychologist, 41,* 1029-1039.

Snyderman, M. & Rothman, R. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137-144.

Spearman, C. (1904). "General intelligence" objectively determined and measured. In J. J. Jenkins, & D. G. Patterson (Eds.), *Studies in individual differences: The search for intelligence.* (pp. 59-74). London: Methuen.

Spector, P. E. (1981). *Research designs.* London: Sage.

Spector, P. E., Fox, S. (2010a). Counterproductive work behavior and organizational citizenship behavior: Are they opposite forms of active behavior? *Applied Psychology: An international Review, 59,* 21-39.

Spector, P. E., Fox, S. (2010b). Theorizing about the deviant citizen: an attributional explanation of the interplay of organizational citizenship and counterproductive work behavior. *Human Resource Management Review, 20,* 132-143.

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2005). The dimensionality of counterproductivity: Art all counterproductive behaviors created equal? *Journal of Vocational Behavior, 68,* 446-460.

Spencer, L. M, & Spencer, S. M. (1993). *Competence at work: Models for superior performance.* New York, NY: Wiley.

Spicer, J. (2005). *Making sense of multivariate data analysis.* Thousand Oaks, CA: Sage.

Spier, P. (2001). *Conviction and sentencing of offenders in New Zealand: 1991 to 2000* (ISSN 1172–0638). Wellington, New Zealand: Ministry of Justice.

Srivastava, S., Guglielmo, S., Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology, 98*, 520-534.

Stankov, L. (1980). Psychometric factors as cognitive tasks: A note on Carroll's "new structure of intellect." *Intelligence,* 4, 65-71.

Stankov, L. (1987). Competing tasks and attentional resources: Exploring the limits of primary-secondary paradigm. *Australian Journal of Psychology,* 39, 123-137.

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

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12,* 35-51.

Stankov, L. (2002). *g*: A diminutive general. In R. J. Sternberg & E. L. Grigorenko (Eds), *The general factor of intelligence: How general is it?* (pp. 19-37). Mahway, NJ Lawrence Erlbaum.

Stankov, L., Boyle, G. J., & Cattell, R. B. (1995). Models and paradigms in personality and intelligence research. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, (pp. 15-43). New York, NY: Plenum Press.

Stankov, L., & Chen, K. (1988). Can we boost fluid and crystallized intelligence? A structural modelling approach. *Australian Journal of Psychology,* 40, 363-376.

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

Stankov, L., Seizova-Cajic, T., & Roberts, R. D. (2001). Tactile and kinaesthetic perceptual processes within the taxonomy of human cognitive abilities. *Intelligence, 29,* 1-29. Abstract.

Stanton Corporation (2004). *A summary of Stanton Survey: New Edition validation*. Technical Report. Charlotte, IL: Stanton.

Stark, S. & Cheryshenko, O. S. (2006). Using modern measurement technology to improve assessment: An introduction to IRT models, concepts, and CAT. Retrieved from http://www.pdf-finder.com/pdf/computerized-adaptive-testing-cat-overview.html.

Stark, S., Chernyehenko, O. S., Chan, K., Lee, W. C., & Drasgow, F. (2001). Effects of the Testing Situation on Item Responding: Cause for Concern. *Journal of Applied Psychology, 86*, 943–953.

State Sector Act 1988 No 20 (as at 01 October 2008), Public Act. Retrieved from: http://legislation.govt.nz/act/public/1988/0020/7.0/DLM129110.html

Staw, B. M., Bell, N. E., & Clausen, J. A. (1986). Dispositional approach to job attitudes: A lifetime longitudinal test. *Administrative Science Quarterly, 31,* 56-77.

Sternberg, R. J. (Ed.) (1982). *Handbook of human intelligence.* Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (1986). *Beyond IQ: A triarchic theory of intelligence.* New York, NY: Cambridge University Press.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* New York, NY: Cambridge University Press.

Sternberg, R. J. (Ed.-in-Chief) (1994). *Encyclopaedia of human intelligence* (Vols. 1 and 2.). New York, NY: Macmillan/Maxwell Canada/Maxwell Macmillan International.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology, 3,* 292-316.

Sternberg, R. J. (2000). Implicit theories of intelligence as exemplar stories of success: Why intelligence test validity is in the eye of the beholder. *Psychology, Public Policy, and Law, 6,* 159-167.

Sternberg, R. J. (2002a). Beyond *g*: The theory of successful intelligence. In R. J. Sternberg & E. L. Grigorenko (Eds), *The general factor of intelligence: How general is it?* (pp. 447-479). Mahwah, NJ: Lawrence Erlbaum.

Sternberg, R. J. (2002b). *The antimony of impact: What makes the great works of psychology great.* New York, NY: Cambridge University Press.

Sternberg, R. J. (2003). A broad view of intelligence: The theory of successful intelligence. *Consulting Journal of Psychology: Practice and Research, 55,* 139-154.

Sternberg, R. J. (2004). North American approaches to intelligence. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 411-444). Cambridge, UK: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (Eds.) (2002). *The general factor of intelligence: How general is it?* Mahwah, N. J.: Lawrence Erlbaum.

Sternberg, R. J., Lautrey, J., & Lubart, T. I. (2003). Where are we in the field of intelligence, how did we get here, and where are we going? In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: International perspectives*, (pp. 3-25). Washington, D.C: American Psychological Association.

Sternberg, R. J., & Wagner, R. K. (Eds.) (1994). *Mind in context: Interactionist perspectives on human intelligence.* Cambridge, UK: Cambridge University Press.

Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84,* 959-968.

Stewart-Williams, S. (2004). *Darwin meets Socrates: Evolutionary psychology and the innate ideas debate.* Unpublished Doctoral Dissertation, Massey University.

St George, R. (1983). Some psychometric properties of the Queensland Test of Cognitive Abilities with New Zealand European and Maori children. *New Zealand Journal of Psychology, 12,* 57-68.

Stoddard, G. D. (1943). *The meaning of intelligence.* New York, NY: Macmillian.

Stone-Romero, E. F., Alvarez, K., & Thompson, L. F. (2009). The construct validity of conceptual and operational definitions of contextual performance and related constructs. *Human Resource Management Review, 19,* 2.

Strand, A. L., & Strand, M. L. (1986). *Validity, reliability, compliance and efficacy studies of the T.A, A.I., E.S., D.A, and P.A.S.S. Surveys*. Oak Brook, IL: Psychological Surveys Corporation.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement, 17,* 277-292.

Stuart-Hamilton, I. (1996). *Dictionary of psychological testing, assessment and treatment.* London: Jessica Kingsley Publishers.

Tabachnick, B. G. & Fidell, L. S. (2001). Principal components and factor analysis. In *Using multivariate statistics* (4th ed., pp. 582 - 633). Needham Heights, MA: Allyn & Bacon.

Taylor, P. J., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organisations and recruitment firms. *New Zealand Journal of Psychology, 31,* 8-18.

Taylor, P. J., Mills, A., & O'Driscoll, M. (1993). Personnel selection methods used by New Zealand organisations and personnel consulting firms. *New Zealand Journal of Psychology, 22,* 19-31.

Talyor, E. K., & Nevis, E. C. (1961). Personnel selection. *Annual Review of Psychology, 12,* 389-412.

Tatsuoka, K. K. (2009). *Cognitive assessment: an introduction to the rule space method.* New York, NY: Routledge.

Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychological Inquiry,* 4, 122-130.

Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology, 30,* 47-54.

Terman, L. M. (1916). *The measurement of intelligence.* Boston, MA: Houghton Mifflin.

Terris, W. (1985). Attitudinal correlates of employee integrity. *Journal of Police and Criminal Psychology, 1,* 60-68.

Terris, W. (1986). *The development and validation of EPI-3.* Park Ridge, IL: London House, Inc.

Terris, W., & Jones, J. W. (1982). Psychological factors related to employees' theft in the convenience store industry. *Psychological Reports, 51,* 1219-1238.

Tetrick, L. E. (1989). An exploratory investigation of response latency in computerized administrations of the Marlow-Crowne social desirability scale. *Personality and Individual Differences, 10,* 1281-1287.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500-517.

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: testing a principle of trait activation. *Journal of Research in Personality, 34,* 397-423.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior, 24,* 335-356.

Thibodeaux, H. F., & Kudisch, J. D. (2003). The relationship between applicant reactions, the likelihood of complaints, and organizational attractiveness. *Journal of Business and Psychology, 18,* 247=257.

Thissen, D. (1983). Timed testing: Ad approach to using item response theory. In D. J. Weiss (ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 179-203). New York: Academic Press.

Thissen, D. (1991). *MULTILOG user's guide (version 6.0).* Mooresville, IN: Scientific Software.

Thomas, J. C. (2004). Overview. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment: Vol. 4, Industrial and organizational assessment,* (pp. 1-3). New Jersey: Wiley.

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

Thorndike, E. L. (1913). *Mental and social measurement.* New York, NY: Teachers College Publications.

Thorndike, E. L. (1927). *The measurement of intelligence.* New York, NY: Bureau of Publications, Teacher's College, Columbia University.

Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research, 20,* 241-254.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometrika Monograph, 1,* 121.

Thurstone, L. L. (1947). *Multiple factor analysis: a development and expansion of the vectors of the mind.* Chicago, IL: Chicago University Press.

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence.* Chicago, IL: University of Chicago Press.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlmen, K., Segall, D. O., & Shephard, W. (2006). Unproctored Internet testing in employment settings. *Personnel Psychology, 59,* 189-225.

Todd, P. M., & Gigerenzer, G. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences, 23,* 727-780.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375-424.

Trapnell, P. D. (1994). openness versus intellect: A lexical left turn. *European Journal of Personality, 8,* 273-290.

Traub, R. (1997). Classical Test Theory in historical perspective. *Educational Measurement, 16,* 8–14.

Treaty of Waitangi Act 1975 No 114 (as at 05 August 2009), Public Act. Retrieved from: http://www.legislation.govt.nz/act/public/1975/0114/16.0/DLM435368.html

Truxillo, D. M., Steiner, D. D., & Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: defining when selection fairness really matters. *International Journal of Selection and Assessment, 12,* 39-53.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings, *Technical Report No.* ASD –TR-61-97, U.S., Air Force, Lackland Air Force Base, TX. *Journal of Personality*, 60, 225-251.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

U.S. Equal Opportunity Employment Commission, U.S. Civil Service Commission, U.S. Department of Labor, U.S. Department of Justice. (1978) Uniform guidelines on employee selection procedures. *Federal Register, 43,* 38295-38309.

van Breukelen, G. J. P. (2005). Psychometric modelling of response speed and accuracy with mixed and conditional regression. *Psychometrika, 70,* 359-376.

van der Linden, W. J. (2007). A hierarchical framework for modelling speed and accuracy on test items. *Psychometrika, 72,* 287-308.

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics, 33,* 5-20.

van der Linden, W. J. (2009). Predictive control of speededness in adaptive testing. *Applied Psychological Measurement, 33,* 25-41. Retrieved from http://apm.sagepub.com/content/33/1/25

van der Linden, W. J., Entink., R. H. K., & Fox, J. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement, 34,* 327-347. Retrieved from http://apm.sagepub.com/content/34/5/327

van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika, 73,* 365-384.

van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika, 68,* 251-265.

Vandenberg, S. G. (1967). The primary mental abilities of South American students: A second comparative study of the generality of a cognitive factor structure. *Multivariate Behavioral Research,* 2, 175-198.

Van der Zee, K. I., Zaal, J. N., & Piekstra, J. (2003). Validation of the multicultural personality questionnaire in the context of personnel selection. *European journal of personality*, 17(1), s77-s100.

Vardi, Y., & Weitz, E. (2004). *Misbehavior in organizations: Theory, research, and management.* Mahwah, NJ: Lawrence Erlbaum.

Varma, S. 2006. Preliminary item statistics using point-biserial correlation and pvalues. Retrieved from http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf.

Vasilopoulos, N. L., Cucina, J. M., & Hayes, T. L. (2005). Using response accuracy and latency scores from an artificial language test to predict grades in a Spanish language course. *International Journal of Selection and Assessment, 13*, 92-95.

Vasilopoulos, N. L., Reilly, R., & Leaman, J. A. (2000). The Influence of Job Familiarity and Impression Management on Self-Report Measure Scale Scores and Response Latencies. *Journal of Applied Psychology, 85*, 50–64.

Verbic, S. (2010). *Establishing the rhythm of responding to test questions*. Online paper submitted to research resources for physics education, Cornell University Library. Retrieved from http://arxiv.org/abs/1005.0714.

Vernon, P. E. (1950). *The structure of human abilities.* New York, NY: Wiley.

Vernon, P. E. (1956). *The measurement of abilities* (2nd ed). London: University of London Press Ltd.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

Vernon, P. E. (1969). *Intelligence and cultural environment.* London: Methuen.

Vinchur, A. J., Schippman, J. S., Switzer, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 81,* 586-597.

Vispoel, W. P., Boo, J., & Bleiler, T. (2001). Computerized and paper and pencil versions of the Rosenberg self-esteem scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement, 61*, 461-474.

Viswesvaran, C. (2002). Absenteeism and measures of job performance: A meta-analysis. *International Journal of Selection and Assessment, 10,* 12-17.

Viswesvaran, C., Deller, J., & Ones, D S. (2007). Personality measures in personnel selection: some new contributions. *International Journal of Selection and Assessment, 15,* 354-358.

Viswesvaran, C., & Ones, D. S. (1997). Review of the Stanton Survey. *Security Journal, 8,* 167-169.

Viswesvaran, C., & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (GMA) in Industrial, Work, and Organizational Psychology. *Human Performance, 15,* 211-231.

Viswesvaran, C., & Ones, D. S. (2004). Importance of perceived personnel selection system fairness determinants: Relations with demographic, personality, and job characteristics. *International Journal of Selection and Assessment, 12*, 172-186.

Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment, 9,* 277-289.

Viteles, M. S. (1930). Psychology in industry. *Psychological Bulletin, 27,* 567-635.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Trans.). Cambridge, MA: Harvard University Press. (Original work published 1930, 1933, 1935).

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Applied Cognitive Psychology, 17,* 755-774.

Waller, N. G., & Ben-Porath, Y. S. (1987). Is it time for clinical psychology to embrace the five-factor model of personality? *American Psychologist,* 42, 887-889.

Wanek, J. E. (1995). The construct of integrity: Item level factor analysis of the dimensions underlying honesty testing and big-five measures of personality. *Dissertation Abstracts International:* UMI No. 9541372.

Wanek, J. E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection and Assessment, 7,* 183-195.

Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology, 56,* 873-894.

Wang, T., & Ni, J. (2007). Speedup analysis for parallel implementation of model of response accuracy and response time in computerized testing. *Lecture Notes in Computer Science, 4489,* 34-41.

Wang, T., & Zhang, J. (2006). Optimal partitioning of testing time: theoretical properties and practical implications. *Psychometrika, 71,* 105-120.

Warr, P., Bartram, D., & Martin, T. (2005). Personality and sales performance: Situational variation and interactions between traits. *International Journal of Selection and Assessment, 13,* 87-91.

Watson, G. (1932). Measures of character and personality. *Psychological Bulletin, 29,* 147-176.

Watson, G. (1933). Character and personality tests. *Psychological Bulletin, 30,* 467-487.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4^{th} ed.). New York: Psychiatric Division, Bellevue Hospital.

Wechsler, D. (1974). *Wechsler intelligence scale for children – revised.* New York: Psychological Corporation.

Weinfurt, K. P. (2003). Multivariate analysis of variance. In L. G. Grimm & P. R. Yarnold, *Reading and understanding multivariate statistics* (8^{th} ed.), (pp. 245-277). Washington, DC: American Psychological Association.

Welch, J., & Welch, S. (2005). *Winning.* New York, NY: HarperCollins Publishers, Inc.

Wenger, M. J. (2005). Models for the statistics and mechanisms of response speed and accuracy. *Psychometrika, 70,* 383-388.

Werner, J. M. (1994). Dimensions that make a difference: Examining the impact of in-role and extra-role behaviors on supervisory ratings. *Journal of Applied Psychology, 79,* 98-107.

Werner, S. H., Jones, J. W., & Steffy, B. D. (1989). The relationship between intelligence, honesty, and theft admissions. *Educational and Psychological Measurement, 49,* 921-927.

West, M. A. (2002). The human team: basic motivations and innovations. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 2: Organizational Psychology), (pp. 270-288). London: Sage.

Wicherts, J. M., Borsboom, D., & Dolan, C. R. (2009). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences,* in press.

Wickett, J. C., & Vernon, P. A. (2000). Replicating the movement time-extraversion link…with a little help from IQ. *Personality and Individual Differences, 28,* 205-215.

Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. *Annual Review of Psychology, 43,* 473-504.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York, NY: McGraw-Hill.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalisability across jobs and performance factors. *Personnel Psychology, 43,* 355-347.

Wise, S. L., & DeMars, C. E. (2005). Examinee motivation in low-stakes assessment: problems and potential solutions. *Educational Assessment, 10,* 1-18.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: the effort-moderated IRT model. *Journal of Educational Measurement, 43,* 19-38.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 16,* 163-183.

Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology, 87*, 164-169.

Wittmann, W. W. (2005). Group differences in intelligence and related measures. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence,* (pp. 223-239). Thousand Oaks: Sage.

Wolfson, S. (2000). Students' estimates of the prevalence of drug use: Evidence for a false consensus effect. *Psychology of Addictive Behaviors, 14,* 295-298.

Wood, P. Q. (2004). *Personality correlates of fluid and crystallized intelligence in the New Zealand workplace*. Unpublished Master's thesis, Massey University.

Wood, P. Q., & Englert, P. (2009). Intelligence Compensation Theory: A critical examination of the negative relationship between conscientiousness and Fluid and Crystallized Intelligence. *Australian and New Zealand Journal of Organisational Psychology, 2,* 19-29.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Woodcock, R. W. (1994). Fluid and crystallized theory of intelligence, measures of. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 452-456). New York: Macmillan.

Woodcock, R. W. (1995). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Woods, S. A., & West, M. A. (2010). *The psychology of work and organisations.* London: Cengage Learning EMEA.

Woolley, R. M., & Hakstian, A. R. (1992). An examination of the construct validity of personality-based and overt measures of integrity. *Educational and Psychological Measurement, 52,* 475-489.

Young, K. (1923). The history of mental testing. *Pedagogical Seminary, 31,* 1-47.

Wuensch, K. L. (2010). *A brief introduction to reliability, validity, and scaling*. Retrieved from http://core.ecu.edu/psyc/wuenschk/MV/FA/Reliability-Validity-Scaling.doc

Yang, J., & Diefendorff, J. M. (2009). The relations of daily counterproductive workplace behaviour with emotions, situational antecedents, and personality moderators: A diary study in Hong Kong. *Personnel Psychology, 62,* 259-295.

Yang, X. (2007). Methods of identifying individual guessers from item response data. *Educational and Psychological Measurement, 67,* 745-764. Retrieved from http://epm.sagepub.com/content/67/5/745

Zeidner, M. (1995). Personality trait correlates of intelligence. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*, (pp. 299-319). New York: Plenum Press.

Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 581-610). Cambridge, England: Cambridge University Press.

Zickar, M. J., & Drasgow, F. (1996). Detecting Faking on a Personality Instrument Using Appropriateness Measurement. *Applied Psychological Measurement, 20*, 71-87.

Zubin, K. (1948). Recent advances in screening the emotionally maladjusted. *Journal of Clinical Psychology, 4,* 56-62.

Appendix A
Items Retained Subsequent to Piloting

Table A1

Retained Integrity Items

Integrity Items
If it would otherwise go to waste, most people would take something from work without asking permission (1)
It is okay to bend company rules when necessary (3)
If I saw someone who looks rich drop a small amount of money I would keep it (4)
Shoplifting from a store is not as bad as stealing from an individual (7)
Sometimes the rules need to be broken to get the job done on time (8)
I have borrowed things from where I have worked without anyone knowing about it (9)
I would keep the money if I were given too much change in a store (10)
Most of us have stolen something when we were young (12)
It is okay to cheat someone who has cheated you first (16)
Supervisors expect employees to try and get away with breaking company rules (17)
Most people would steal something if they knew they would not get caught (18)
Taking something from work without permission is not always “stealing” (19)
If an employee’s hard work is not recognised they will ignore company rules (21)
I would help a friend get back at a bad boss (24)
Employees caught stealing things of little value should be given a second chance (30)
Sometimes employees have to stretch the rules to get the job done (31)
Employees often steal because they are not treated fairly by their boss (32)
If I were accidentally given too much change in a store, I would probably keep it (37)
There is nothing that could tempt me to steal (38)
Some people have good reasons for stealing (39)

Note. The number within parenthesis denotes the original item number as piloted.

Table A2

Retained Conscientiousness Items

Conscientiousness Items
I like to leave my options open instead of planning everything in advance (1)
I am a well organised person (3)
I often leave things as long as possible before trying to get them done (4)
I complete things just before they need to be done (10)
I notice when things are out of place (11)
I often go with what I feel rather than thinking things through (12)
I often can’t find where I have put things (15)
I work in an orderly way towards a clear set of goals (16)
I like to take life as it comes rather than plan too far ahead (19)
I like making detailed plans (23)
People have sometimes said that I am disorganised (27)
People have told me I have a lot of self-discipline (29)
I tend to put things neatly away (30)
I am a person with a lot of self-discipline (31)
I am more interested in enjoying life every day than worrying about the future (34)
I like things to be in their proper place (35)
I make sure I keep records in order to avoid forgetting important dates or details (42)
I prefer to take things as they come rather than having a strict routine (45)
I am very competitive (48)
I work hard to achieve the goals I have set myself (50)

Note. The number within parenthesis denotes the original item number as piloted.

Table A3

Retained Neuroticism Items

Neuroticism Items
I often worry about things (5)
I takes a lot to get me angry (6)
I often feel unable to cope with problems (9)
My mood changes a lot (11)
I often feel uncomfortable when I am around people I do not know (13)
I often feel nervous (15)
It does not take much to make me angry (16)
I sometimes feel guilty (18)
I am quick to anger (23)
I sometimes wonder if people really like me (24)
I sometimes worry about saying something embarrassing when around other people (25)
I often think about my past mistakes (30)
I get nervous before important things I have to do (32)
I am often discouraged when people make negative comments (34)
I find myself replaying things that happen again and again in my mind (39)
My lack of self-confidence means I do not do as well as I could in social situations (40)
People have told me I am moody (41)
I sometimes doubt myself (43)
I often feel worried about how things will go in the future (44)
I get upset by the mean things other people say (47)

Note. The number within parenthesis denotes the original item number as piloted.

Table A4

Retained Dissimulation Items

Dissimulation Items
I have never broken a promise (1)
I never talk about people behind their back (3)
I never tell lies (5)
I never judge people before I know them (7)
I always do the right thing (8)
Everything I do is interesting (9)
I always do exactly what is expected of me (10)
I have never failed to reach a personal goal I have set for myself (11)
I like everyone I meet (17)
I have always done what my parents have asked me to do (18)

Note. The number within parenthesis denotes the original item number as piloted.

Table A5

Retained Crystallised Intelligence Items

Crystallised Items

The meaning of *prudent* is closest to? (item 1)

Guarded	Shy	Educated	Reserved	Sensible

The meaning of *mixed* is closest to? (item 3)

Altered	Similar	Parallel	Sundry	Different

The meaning of *anticipated* is closest to? (item 6)

Expected	Normal	Accepted	Predictable	Customary

Which of the following is the odd one out? (item 14)

Goalie	Football	Wrestling	Tennis	Golf

Which of the following is the odd one out? (item 15)

Job	Worker	Career	Employment	Occupation

Which of the following is the odd one out? (item 16)

Era	Time	Year	Period	Age

Which of the following is the odd one out? (item 18)

Elephant	Whale	Tiger	Dolphin	Shark

The meaning of *gregarious* is furthest from? (item 20)

Friendly	Grumpy	Unhappy	Unsocial	Unstable

The meaning of *ignorant* is furthest from? (item 22)

Sure	Aware	Attentive	Awake	Impolite

The meaning of *organised* is furthest from? (item 24)

Retained	Tidy	Dishevelled	Obtuse	Discarded

The meaning of *income* is furthest from? (item 25)

Payments	Expenditure	Bills	Statements	Tax

An *eddy* is a? (item 26)

Small Stream	Cave Dweller	Whirlpool	Small Tree	Small Fossil

A *curmudgeon* is a? (item 27)

Ladle	Grump	Pigeon Species	Weapon	Spice

An *invoice* is a? (item 29)

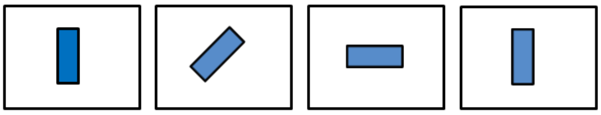
Demand	Schedule	Ticket	Receipt	Bill

Note. The number within parenthesis denotes the original item number as piloted.

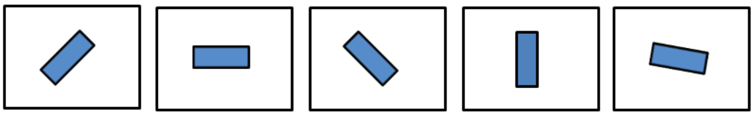
Table A6

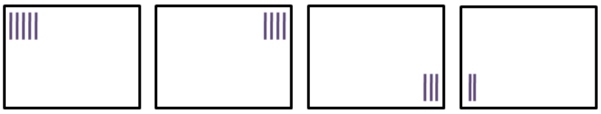
Retained Fluid Intelligence Items

Fluid Items

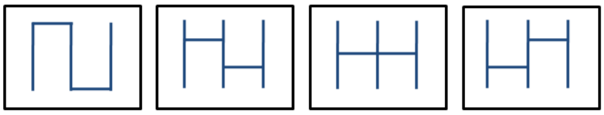


What comes next?
(item 1)

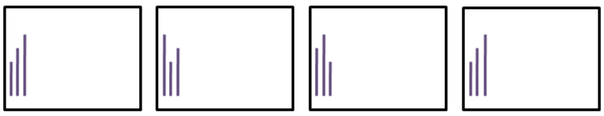




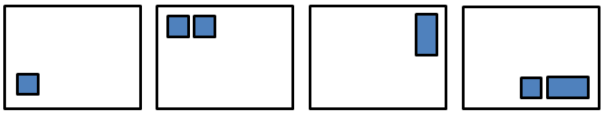
What comes next?
(item 2)



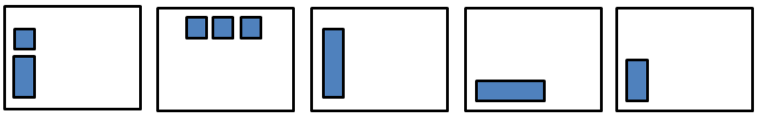
What comes next?
(item 4)

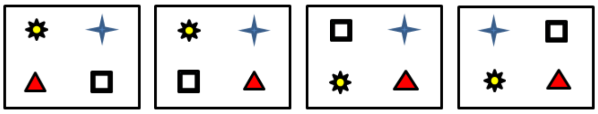


What comes next?
(item 8)

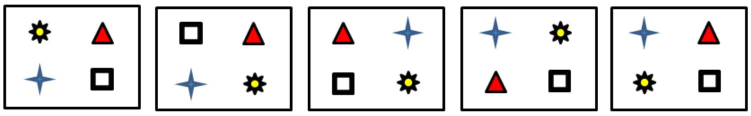


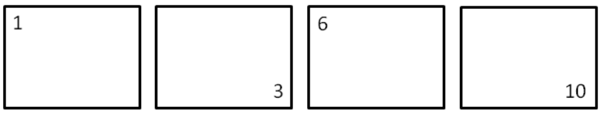
What comes next?
(item 9)



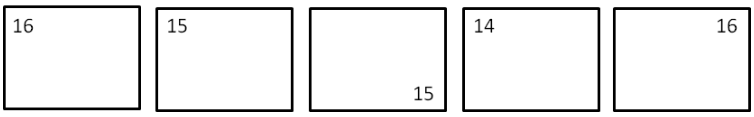


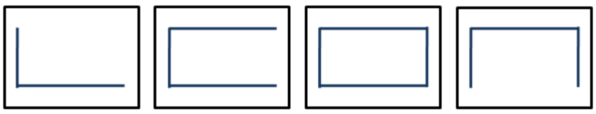
What comes next?
(item 10)





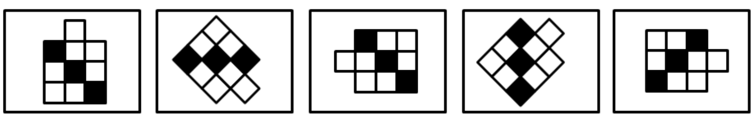
What comes next?
(item 11)



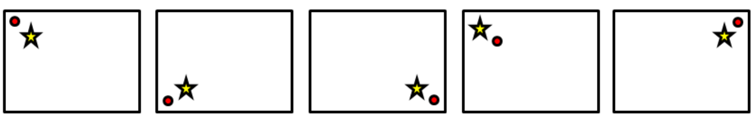


What comes next?
(item 12)

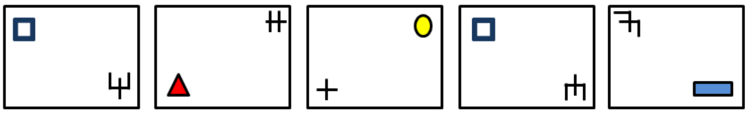
Which is the odd one out? (item 16)



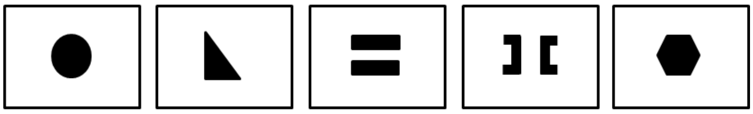
Which is the odd one out? (item 19)



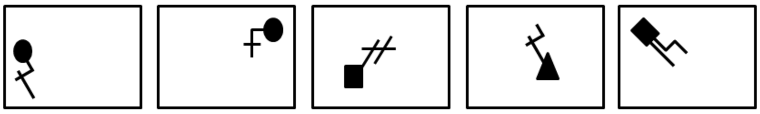
Which is the odd one out? (item 21)



Which is the odd one out? (item 22)



Which is the odd one out? (item 23)



Which is the odd one out? (item 24)

Which is the odd one out? (item 25)

2091420 is to tint as 1015914 is to? (item 26)

Lead	Join	Loop	Crane	Joke

Note. The number within parenthesis denotes the original item number as piloted.

Appendix B

Participant Information and Consent

Dear Participant

My name is Paul Wood. I work for a company called OPRA in the area of workplace psychology. I am a registered doctoral student at Massey University. I invite you on the following adventure as part of my doctoral research. By clicking "agree" at the bottom of this page you have the opportunity to become familiar with the types of questions encountered when applying for employment or promotion. Becoming familiar with such questions is likely to increase your score on ability tests and allow you to present yourself accurately on personality assessments. At the end of the three assessments you will also have the opportunity to request a summary of the research and make any comments about the assessments.

The research aims at improving procedures of assessment when applicants are assessed for employment. This will increase people's chances of getting the good jobs they want and are suited to. This will increase workplace diversity by reducing the unfair treatment that some New Zealanders face when poor selection methods are used. OPRA will be offering the final assessment designed as a result of this research.

The study involves three mini-assessments. It will take approximately 15 to 20 minutes to complete each of them. All data collected are private and anonymous. Your name will not be kept in any data file. Any information you provide is strictly confidential. Information will only be published as statistical summaries without names. Your name and individual results will never be associated in any way with results published – including articles or presentations.

It is entirely your choice whether you choose to take part in this study. You are free to change your mind about taking part and quit at any time. When the study is complete, a summary of findings will be available at www.opragroup.com

Please feel free to ask any questions you may have. My supervisor for this research is Dr G. Haberman (Massey University). He can be contacted at g.m.habermann@massey.ac.nz*, or (06) 3569099. I can be contacted at* paul.wood@opragroup.com*, (09) 358 3233, or 021 972 976.*

Please click "I agree" and proceed to practice questions if all of the below are true:

- *I have read the preceding information and understand the details of the study.*
- *I understand that I may change my mind about participation and log off at any time.*
- *I understand that my responses will be anonymously stored and used for analysis.*
- *I agree to participate in this study under the conditions set out in the preceding information.*
- Agree Disagree

This project has been reviewed and approved by the Massey University Human Ethics Committee: Southern A, Application 08/43. If you have any concerns about the conduct of this research, please contact Professor John O'Neill, Chair, Massey University Human Ethics Committee: Southern A, telephone 06 350 5799 x 8771, email humanethicsoutha@massey.ac.nz

This project is made up of three mini-assessments. The first assessment looks at your understanding of language. The second assessment looks at your ability to solve new problems. Both of these assessments are designed to be challenging. This is so it is possible to see differences between people of all levels of ability. Do not expect to get all of the questions correct. You will get points for answering questions correctly. You will also get points for how quickly you correctly answer questions. If you take longer than 40 seconds to answer a question you will be asked to have a guess and move on. You do not lose points for guessing. If you have still not answered after 50 seconds you will automatically move to the next question. Please complete all three mini-assessments in one sitting.

The third assessment looks at your personality. It looks at how you are likely to behave at work. It assesses your attitudes, beliefs, and how you prefer to behave. More information is provided on this assessment after the first two assessments are complete.

How to answer

To answer a question, simply click on the answer you choose. When you click, you will automatically proceed to the next question.

Language Ability

This assessment will ask you questions about the meaning of language. The quicker you can correctly answer each question the more points you will get. If you have not answered a question after 40 seconds you will be asked to have a guess. You do not lose points for guessing. If you have not answered after 50 seconds you will automatically move to the next question. You cannot go back to earlier questions once you have answered them or been moved on. There are a total of 30 questions to answer. Some questions may have more than one possible answer, but you are asked to find the best out of available answers. Use your mouse or pointer to click on your choice of answer and move to the next question. Please answer each question as quickly and as accurately as you can.

Problem Solving Ability

The questions in this assessment will look at your ability to quickly identify patterns and sequences and otherwise solve problems. The quicker you can correctly answer each question the more points you will get. If you have not answered a question after 40 seconds you will be asked to have a guess. You do not lose points for guessing. If you have not answered after 50 seconds you will automatically move to the next question. You cannot go back to earlier questions once you have answered them or been moved on. There are a total of 30 questions to answer. Some questions may have more than one possible answer, but you are asked to find the best out of available answers. Use your mouse or pointer to click on your choice of answer and move to the next question. Please answer each question as quickly and as accurately as you can.

Personality Instructions

These questions look at your opinions and how you prefer to behave. Please use the rating scale to describe how accurately each statement describes ***you and your views****. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. If you have not answered a question after 40 seconds you will be asked to make a response. If you have not answered after 50 seconds you will automatically move to the next question. You cannot go back to earlier questions once you have answered them or been moved on. There are a total of 160 questions to answer. You may not feel certain about your answers to many of the questions. Be spontaneous and go with the first answer that comes to mind.*

Appendix C:
Oblique Factor Rotations

Factor Rotation of Extracted Cognitive Ability Factors

Table C1 shows very similar results for both PAF and ML oblique rotations. Other than slight variations in factor loadings, the factor loading of .304 on crystallised item six (PAF) that was absent from ML was the most notable difference. The pattern matrices both suggested that factor one is defined by fluid intelligence item loadings. The most important items for an understanding of this factor appear to be items 19, 22, 23, and 26. Each of these items has important loadings of practical significance on factor one. Items 27 and 24 have the strongest loading on factor two. These are identifying the antonym items. The next most relevant crystallised items for factor two are items 22 and 18. These items are antonym and odd-man-out respectively. Crystallised intelligence item 19 loaded across both factors.

Table C2 suggest that there is considerable similarity between the pattern and structure matrices. The most noticeable differences appears to be a significant loading on factor two for fluid item 20 and some minor variance in the strength of factor loadings.

The component correlation matrices in Table C3 show small relationships between the two factors extracted. This is consistent across both PAF and ML methodology. On this basis orthogonal rotation is justified. It also means no second-order factor rotations are recommended (Child, 2006).

Table C1

Cognitive Ability Pattern Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

	PAF Pattern Matrix		ML Pattern Matrix	
Items	1	2	1	2
Crystallised				
1		.36		.36
3				
6				
14				
15				
16				
18		.36		.38
19	.36	.31	.36	.32
20		.30		
22		.36		.35
24		.46		.44
25				
26				
27		.47		.49
29				
Fluid				
1	.40		.39	
2	.30		.31	
4	.44		.44	
8				
9				
10	.48		.48	
11	.44		.45	
12	.45		.45	
16				
19	.64		.64	
20				
21	.36		.35	
22	.59		.59	
23	.51		.51	
24	.45		.44	
25	.38		.38	
26	.53		.53	

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 6 iterations. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 6 iterations.

Table C2

Cognitive Ability Structure Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

	PAF Structure Matrix		ML Structure Matrix	
Items	1	2	1	2
Crystallised				
1		.379		.375
3				
6		.304		
14	.325		.317	
15				
16				
18		.408		.421
19	.466	.426	.464	.442
20				
22		.340		.328
24		.430		.412
25				
26				
27		.449		.460
29				
Fluid				
1	.368		.369	
2	.336		.340	
4	.465		.468	
8	.326		.315	
9	.301		.301	
10	.442		.446	
11	.406		.404	
12	.445		.445	
16	.326		.328	
19	.647		.651	
20	.385	.370	.385	.367
21	.392		.389	
22	.589		.595	
23	.502		.500	
24	.449		.444	
25	.391		.394	
26	.516		.513	

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 6 iterations. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 6 iterations.

Table C3

Cognitive Ability Principal Axis Factoring and Maximum Likelihood Factor Correlation Matrices

PAF Factor Correlation Matrix		ML Factor Correlation Matrix	
1	2	1	2
1.000	.318	1.000	.320
.318	1.000	.320	1.000

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization.

Factor Rotation of Extracted Personality Factors

As with ability items, further factor analysis with factor rotation was undertaken to facilitate factor interpretation and item retention. Table C4 shows that both PAF and ML oblique rotation produced very similar results. Other than slight variations in factor loadings, the factor loadings of .304 and .318 on neuroticism items 10 and 12 respectively ML that is absent from PAF is the most notable difference. The pattern matrices both suggest that factor one is defined by neuroticism item loadings. The most important items for an understanding of this factor appear to be items 9, 23, 11, 16, 32, 35, and 41. Factor two appears most clearly related to conscientiousness items. Items 3, 31, 4, 16, 23, 27, 42, and 50 have the strongest loading on factor two. Factor three is most clearly defined by integrity scale items. The most relevant items are 1, 3, 10, 18, and 37. Factor four loadings come exclusively from dissimulation items. The strongest loadings come from items 1, 8, 12, and 18. The fifth factor is primarily comprised of negative factor loadings from neuroticism items. The most important loadings appear to come from items 43, 4, 24, 33, 39, 40, and 44.

Table C5 suggests that there is considerable similarity between the pattern and structure matrices for personality factors. The most noticeable differences appears to be less unique loadings amongst factors and a less precise specific collection of factor loadings on the fifth factor.

As with those of the ability items, the component correlation matrices in Table C6 show small relationships between the two factors extracted. This is consistent across both PAF and ML methodology. On this basis orthogonal rotation was justified (Thompson, 2004). Again, it also means no second-order factor rotations are recommended (Child, 2006).

Table C4

Personality Pattern Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

	PAF Pattern Matrix					ML Pattern Matrix				
Items	1	2	3	4	5	1	2	3	4	5
Con										
1		.393					.365			
3		.704					.715			
4		.502					.499			
5										
8					.340					.334
10		.437					.425			
11		.404					.400			
12										
15		.368					.365			
16		.595					.586			
19		.406					.385			
23		.527					.532			
24										
26										
27		.532					.544			
28										
29		.454					.468			
30		.415					.415			
31		.611					.618			
32										
34		.333					.316			
35		.468					.469			
36										
41										
42		.540					.551			
45		.441					.419			
46										
48		.382					.383			
50		.559					.563			
N										
2					-.393					-.398
3										
5					-.587					-.587
6	.381					.388				
7					-.398					-.399
8					-.384					-.380
9	.559					.596				
10						.304				
11	.434					.450				
12					-.334	.318				-.324
13					-.503					-.495
14	.339		.333			.349		.338		
15					-.434					-.428
16	.482					.448				
17					-.436					-.439
18					-.489					-.484
19	.357					.343				
20	.394					.414				
22		-.352					-.344			
23	.522					.508				
24					-.575					-.573
25					-.478					-.478
26			.360					.371		
28					-.363					-.372
30					-.403					-.410
31										
32					-.514					-.527
33	.422					.422				

34					-.481					-.481
35	.455					.486				
36	.340		.311			.337		.315		
38										
39					-.570					-.578
40					-.509					-.496
41	.431					.424				
42										
43					-.649					-.650
44					-.504					-.506
46	.336					.339				
47					-.487					-.486
49			.380					.380		
50					-.317					-.303
In										
1			.527					.524		
3			.510					.503		
4			.315					.311		
7			.474					.469		
8			.346					.349		
9			.441					.445		
10			.564					.581		
11										
12			.452					.444		
16			.418					.420		
17			.409					.408		
18			.541					.537		
19			.417					.414		
21			.494					.488		
24			.372					.369		
26			.314					.307		
30			.338					.336		
31			.410					.406		
32			.430					.421		
35										
37			.523					.545		
38			.456					.457		
39			.423					.408		
40										
D										
1				.453					.440	
3										
5				.350					.337	
6										
7				.399					.378	
8				.478					.479	
9				.354					.341	
10				.487					.497	
11				.373					.360	
12		.331		.457			.314		.473	
13		.569					.571			
15				.329	.338				.334	.338
16		.319		.316			.311		.324	
17				.326					.314	
18				.474					.487	

Note. “Con” = conscientiousness items, “N” = neuroticism items, “In” = integrity items, and “D” = dissimulation items. Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 19 iterations. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 19 iterations.

Table C5

Personality Structure Matrices for Principal Axis Factoring and Maximum Likelihood Extraction

		PAF Structure Matrix					ML Structure Matrix			
Items	1	2	3	4	5	1	2	3	4	5
Con										
1		.392					.374			
3		.717					.726			
4		.545	-.348				.544	-.353		
5		.321					.314			
8		.329	-.345		.430		.328	-.342		.426
10		.439					.432			
11		.400					.394			
12		.354					.341			
15		.407					.407			
16		.643					.637			
19		.417					.402			
23		.535					.540			
24										
26		.346			.313		.334			.300
27		.573					.584			
28										
29		.477					.489			
30		.445					.445			
31		.623					.629			
32										
34		.327					.315			
35		.453					.453			
36		.351	-.338	.387			.348	-.341	.380	
41		.317					.308			
42		.551					.560			
45		.463					.449			
46										
48		.328					.328			
50		.573					.579			
N										
2	.334				-.488	.329				-.491
3			.333					.337		
5	.320				-.620	.325				-.621
6	.431					.434				
7	.325				-.453	.327				-.455
8					-.363					-.361
9	.627				-.340	.656				-.328
10	.336					.348				
11	.548		.307		-.372	.561		.306		-.367
12	.421				-.441	.436				-.435
13					-.490					-.484
14	.473		.468		-.315	.483		.472		-.311
15					-.468					-.465
16	.498					.468				
17					-.451					-.452
18					-.540					-.537
19	.454			-.348		.442			-.358	
20	.506				-.432	.523				-.423
22	.330	-.348				.347	-.343			
23	.554					.541				
24					-.600					-.598
25					-.465					-.465
26	.330		.430			.319		.437		
28					-.364					-.372
30	.338				-.466	.330				-.470
31	.322				-.303	.326				-.302
32					-.513					-.520
33	.530				-.416	.531				-.417

34					-.520					-.521
35	.562				-.449	.587				-.440
36	.419		.408	-.326		.415		.410	-.324	
38			.316		-.325			.315		-.326
39					-.601					-.606
40					-.493					-.485
41	.522				-.358	.516				-.355
42					-.360					-.362
43					-.650					-.651
44	.348				-.570	.351				-.573
46	.451		.309	-.340	-.354	.453		.313	-.350	-.351
47					-.496					-.496
49	.336		.452			.330		.451		
50					-.349					-.341
In										
1			.556					.554		
3			.561					.556		
4			.333					.330		
7			.517					.514		
8			.395					.396		
9			.461					.464		
10			.567					.578		
11			.308					.312		
12			.450					.444		
16			.431					.432		
17			.412					.412		
18			.566					.562		
19			.414					.411		
21			.474					.469		
24			.405					.402		
26			.317					.311		
30			.332					.330		
31			.418					.416		
32			.437					.431		
35			.303					.304		
37			.517					.532		
38			.478					.479		
39			.397					.386		
40			.321					.320		
D										
1				.473					.462	
3				.332					.318	
5				.388					.377	
6										
7				.427					.410	
8				.511					.512	
9				.400					.389	
10				.481					.491	
11				.416					.404	
12		.405		.511			.393		.525	
13		.620		.348			.623		.350	
15				.398	.376				.402	.376
16		.420		.408			.417		.416	
17				.314					.303	
18				.448					.458	

Note. “Con” = conscientiousness items, “N” = neuroticism items, “In” = integrity items, and “D” = dissimulation items. Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization.

Table C6

Personality Principal Axis Factoring and Maximum Likelihood Factor Correlation Matrices

	PAF Correlation Matrix					ML Correlation Matrix				
Factors	1	2	3	4	5	1	2	3	4	5
1	1.000	-.151	.225	-.114	-.303	1.000	-.159	.229	-.113	-.306
2	-.151	1.000	-.246	.222	.110	-.159	1.000	-.246	.224	.104
3	.225	-.246	1.000	-.178	-.203	.229	-.246	1.000	-.181	-.199
4	-.114	.222	-.178	1.000	.131	-.113	.224	-.181	1.000	.126
5	-.303	.110	-.203	.131	1.000	-.306	.104	-.199	.126	1.000

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization.

Appendix D:

Item Response Theory Assumptions and Item Characteristic Curves

Assumption Testing:

Table D1 details the model-data fit of global and individual results for crystallised and fluid intelligence items. As can be seen from Table D1, the chi-squares ranged from .183 to 1.019 for crystallised items and .137 to 2.059 for fluid items. Global chi-squares were a non-significant 6.350 and 14.208 respectively.

Table D1

IRT Model-Data Fit for Crystallised and Fluid Intelligence Items

Item	Chi-square	df	P-Value	Item	Chi-square	df	P-Value
1	0.29	10	1.00	1	0.33	10	1.00
3	0.18	10	1.00	2	1.69	10	0.99
6	0.76	10	1.00	4	0.45	10	1.00
14	0.37	10	1.00	8	0.13	10	1.00
15	0.49	10	1.00	9	1.54	10	0.99
16	0.21	10	1.00	10	2.05	10	0.99
18	0.60	10	1.00	11	0.89	10	1.00
20	0.41	10	1.00	12	0.60	10	1.00
22	0.32	10	1.00	16	0.41	10	1.00
24	0.53	10	1.00	19	0.55	10	1.00
25	0.50	10	1.00	21	0.16	10	1.00
26	0.21	10	1.00	22	0.15	10	1.00
27	0.21	10	1.00	23	0.63	10	1.00
29	1.01	10	1.00	24	0.91	10	1.00
				25	1.90	10	0.99
				26	1.19	10	1.00
Global	5.690	140	1.00	*Global*	12.213	160	1.00

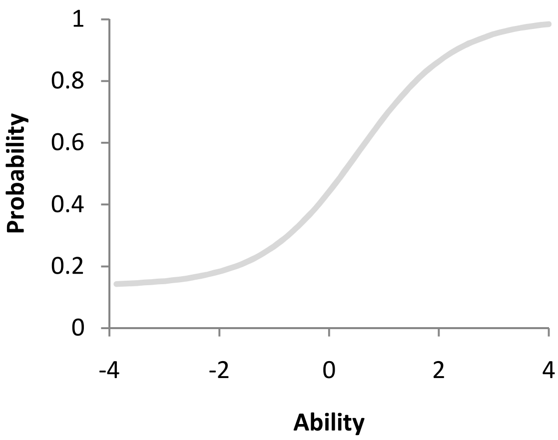


Figure D1. ICC of Gc Item 1

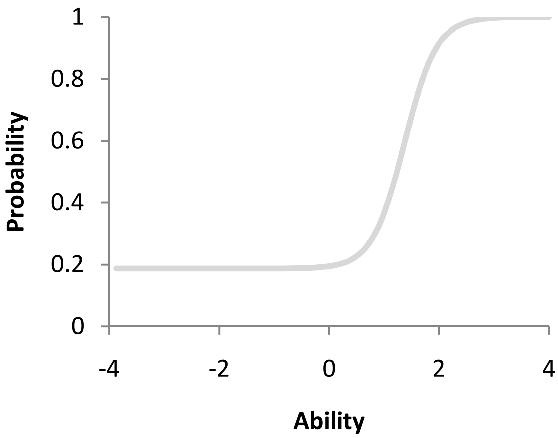


Figure D2. ICC of Gc Item 3

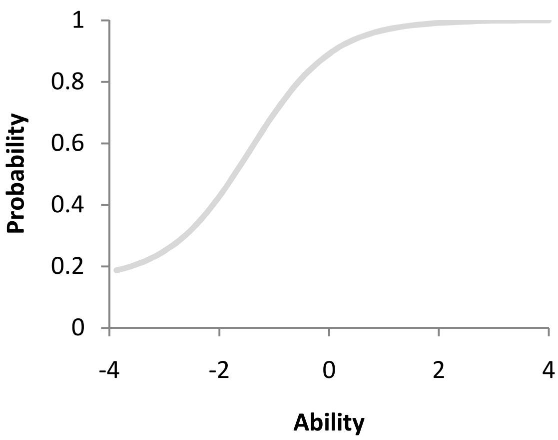


Figure D3. ICC of Gc Item 6

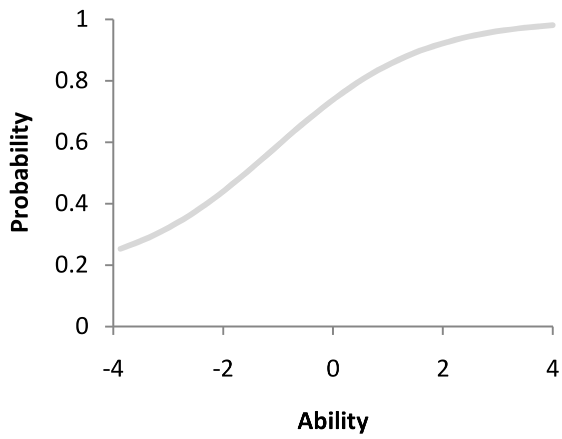


Figure D4. ICC of Gc Item 14

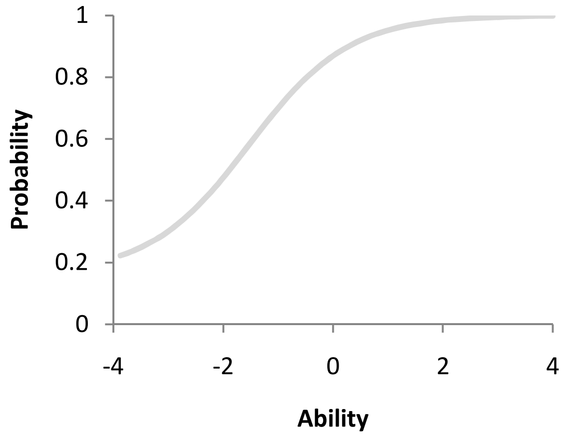


Figure D5. ICC of Gc Item 15

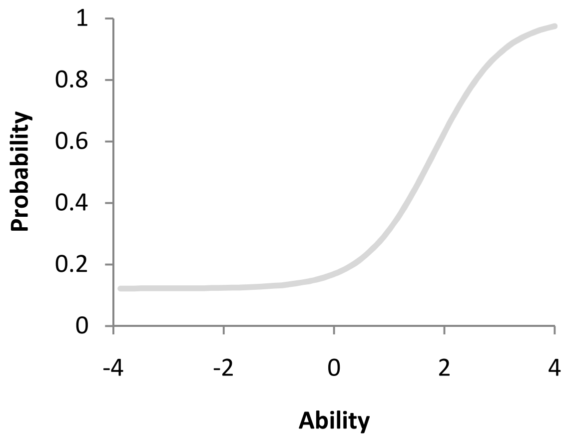


Figure D6. ICC of Gc Item 16

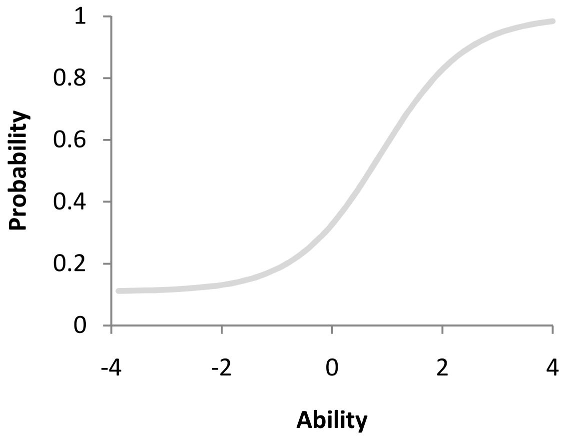


Figure D7. ICC of Gc Item 18

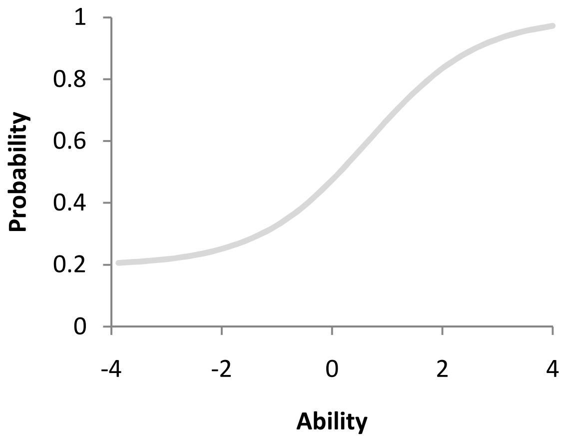


Figure D8. ICC of Gc Item 20

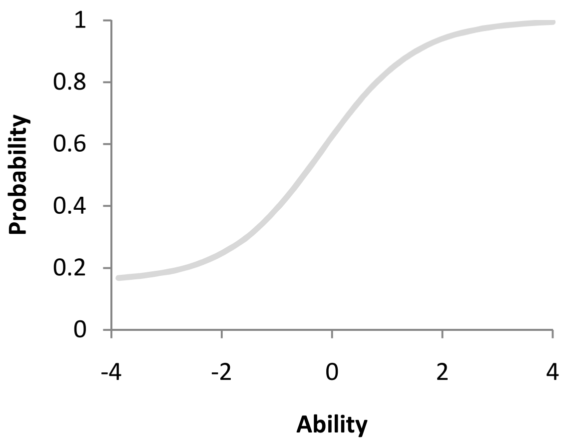


Figure D9. ICC of Gc Item 22

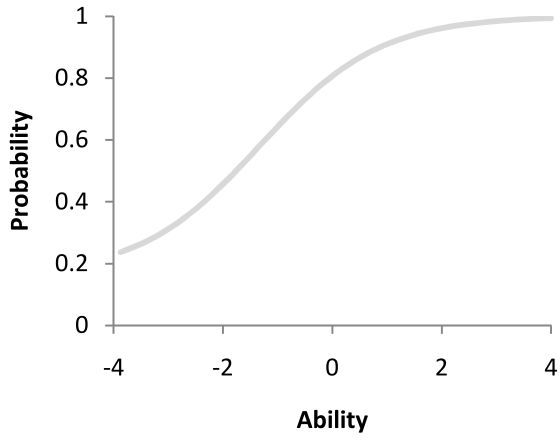


Figure D10. ICC of Gc Item 24

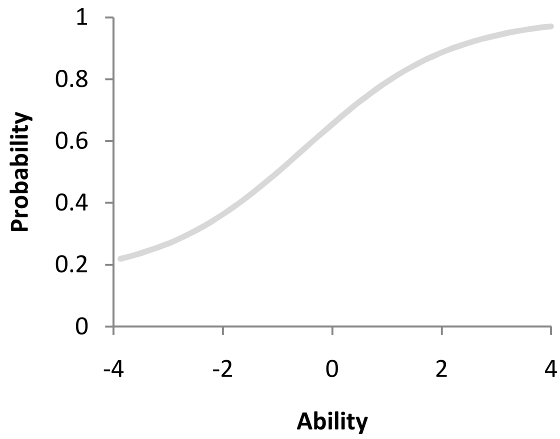


Figure D11. ICC of Gc Item 25

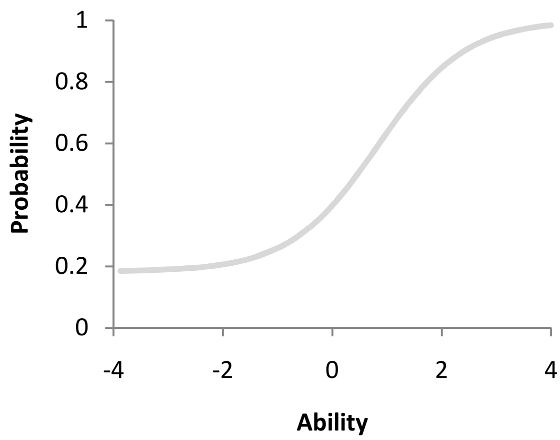


Figure D12. ICC of Gc Item 26

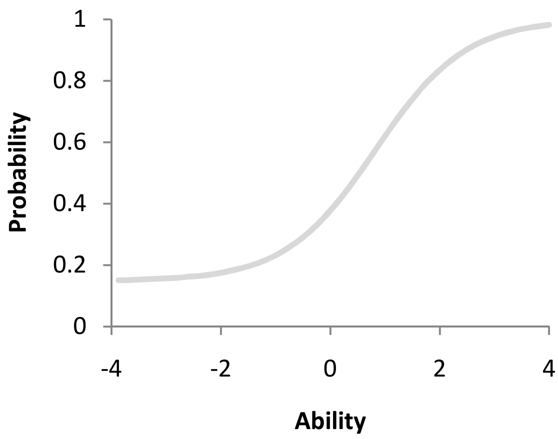


Figure D13. ICC of Gc Item 27

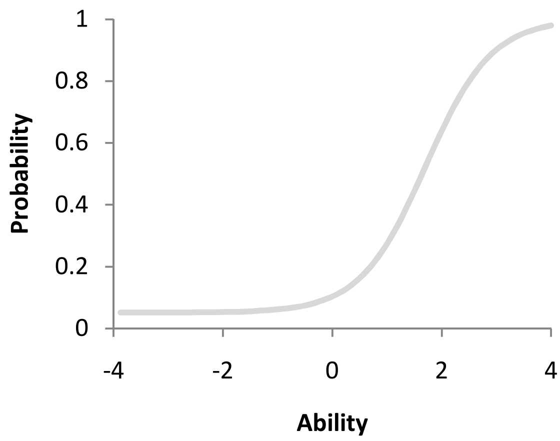


Figure D16. ICC of Gf Item 2

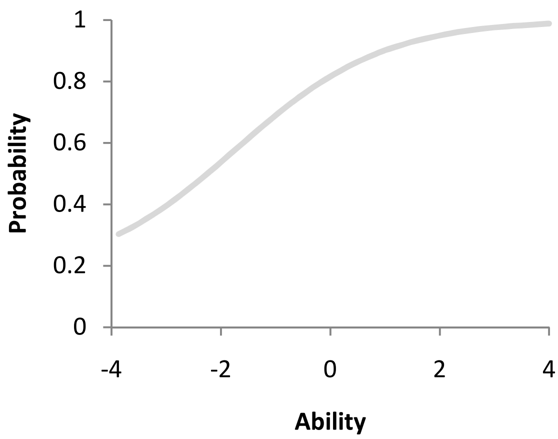


Figure D14. ICC of Gc Item 29

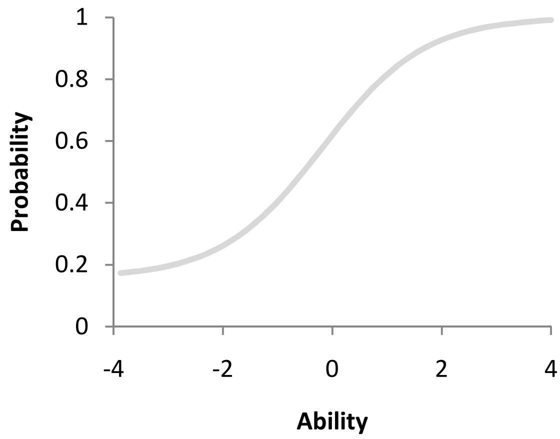


Figure D17. ICC of Gf Item 4

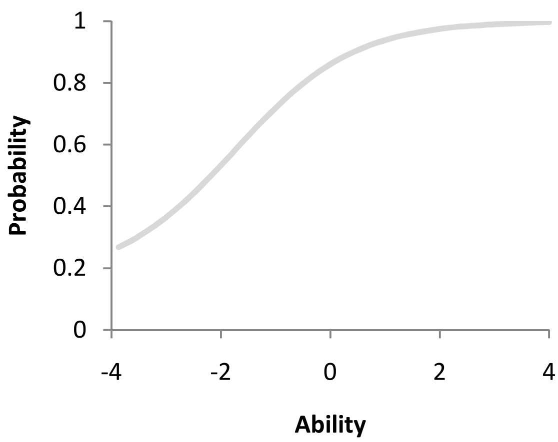


Figure D15. ICC of Gf Item 1

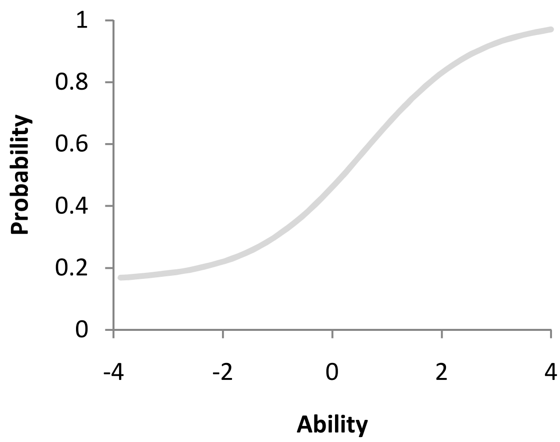


Figure D18. ICC of Gf Item 8

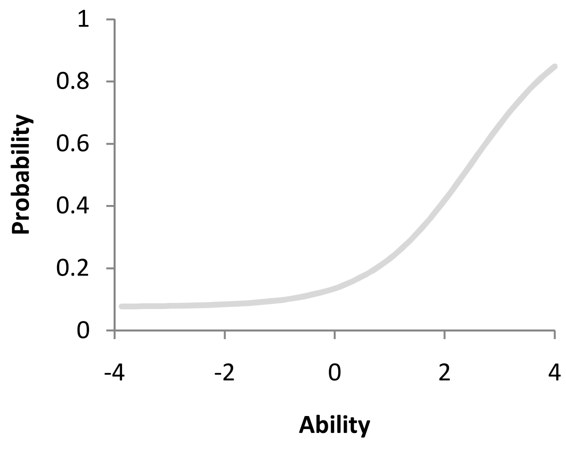


Figure D19. ICC of Gf Item 9

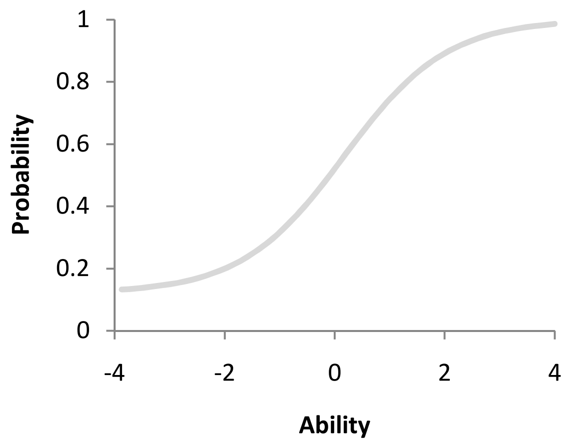


Figure D20. ICC of Gf Item 10

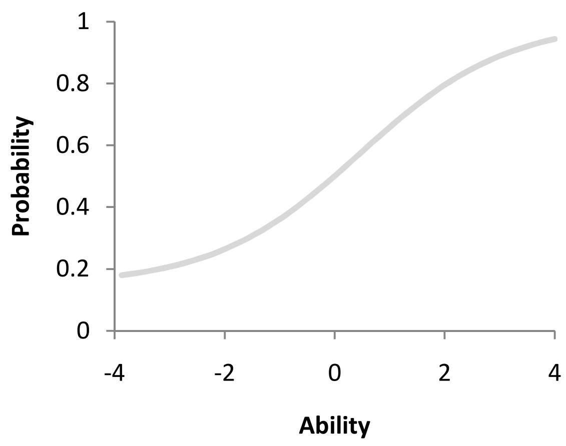


Figure D21. ICC of Gf Item 11

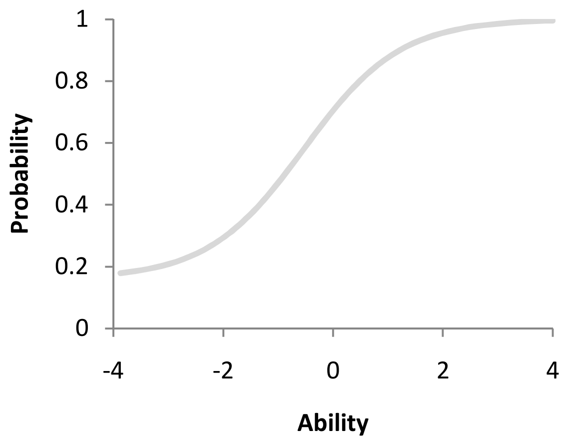


Figure D22. ICC of Gf Item 12

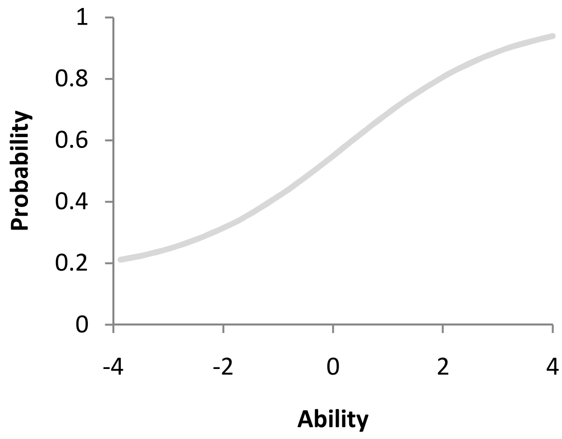


Figure D23. ICC of Gf Item 16

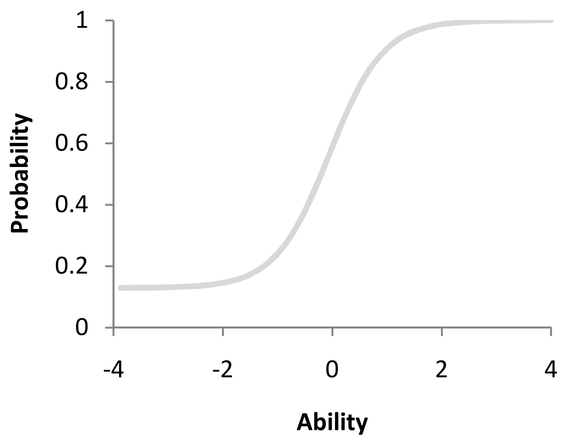


Figure D24. ICC of Gf Item 19

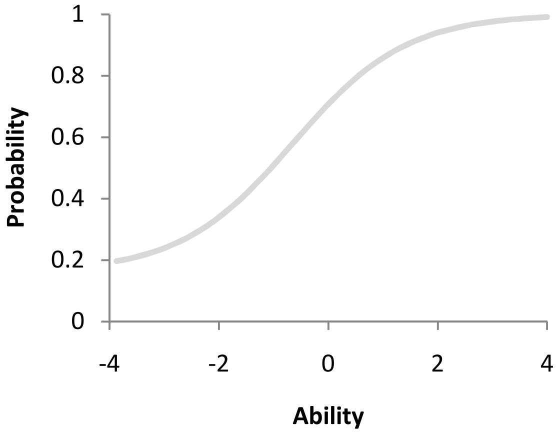


Figure D25. ICC of Gf Item 20

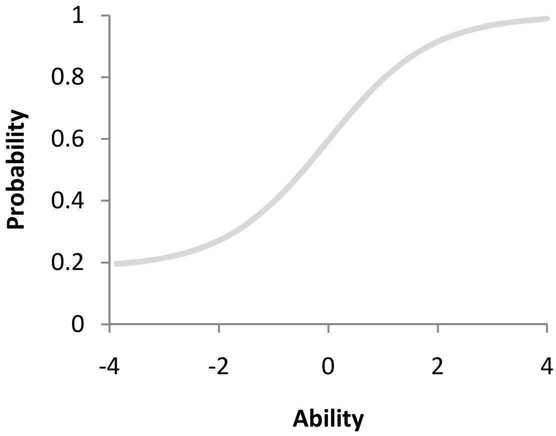


Figure D26. ICC of Gf Item 21

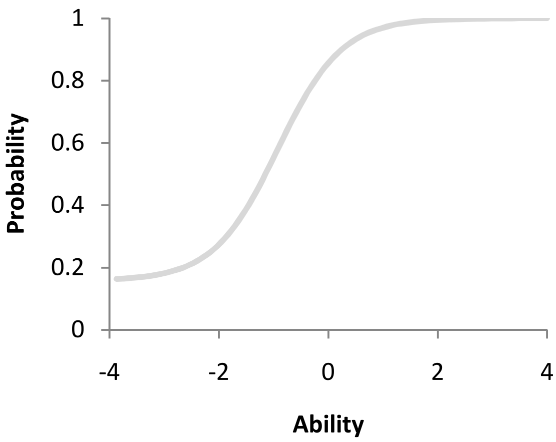


Figure D27. ICC of Gf Item 22

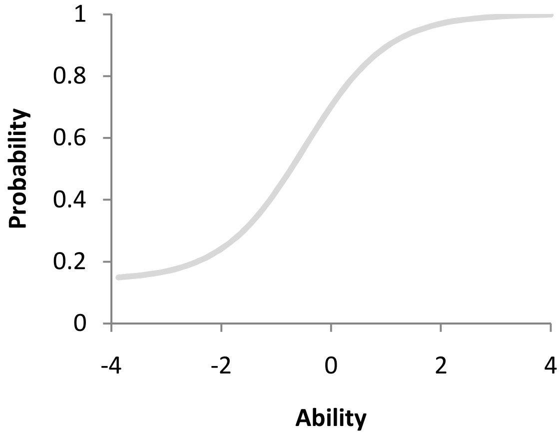


Figure D28. ICC of Gf Item 23

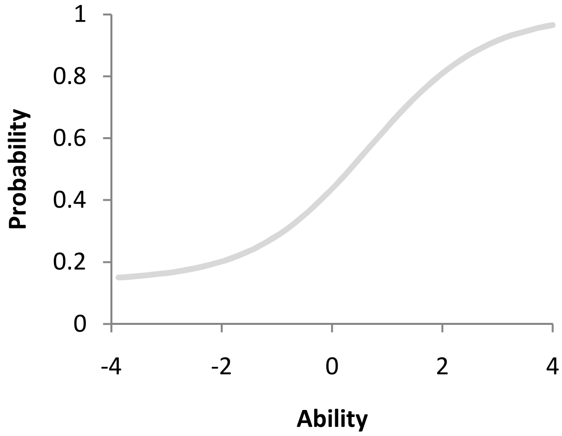


Figure D29. ICC of Gf Item 24

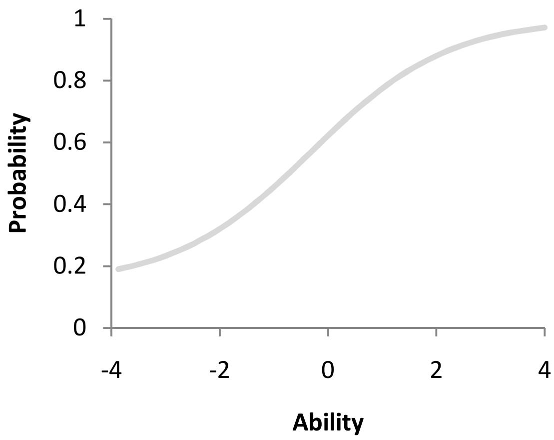
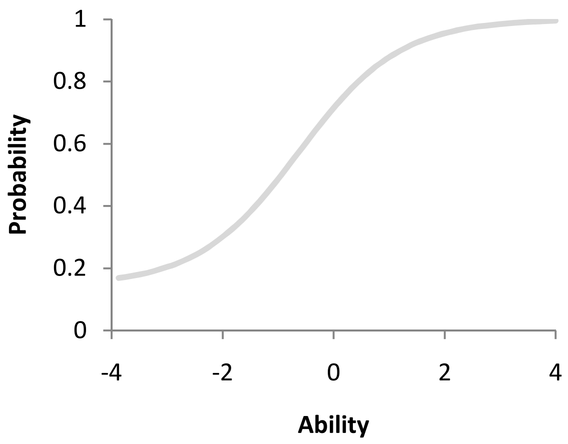


Figure D30. ICC of Gf Item 25



FigureD31. ICC of Gf Item 26