Crystallised and Fluid Ability Change Across Age and a Psychometric Evaluation of the GRT2: A Cross-Sectional Analysis

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

Masters of Arts

in

Psychology

At Massey University, Auckland,

New Zealand.

Andrew James McInnes

2011

Abstract

This paper critically evaluated many psychometric properties of the General Reasoning Test (GRT2) and examined age changes in Crystallised and Fluid ability across the 16- to 58-year age range, as represented by the Verbal and Abstract GRT2 scales, respectively. Respondent data came from a large New Zealand-based archived group (*N*=5075) of individuals whom had completed the assessment as a component of general-level pre-employment information gathering. The psychometric evaluation found the GRT2 to possess adequate internal consistency, and a sub-adequate item-difficulty distribution, for both the Abstract and Verbal scales. A cross-sectional analysis of ability suggested a significant linear decline in Fluid intelligence and non-significant change in Crystallised intelligence, across age. Furthermore, heterogeneity of Fluid ability variance appeared to increase significantly across age, whereas Crystallised ability variance did not. Results were interpreted in the context of Horn's theory, especially with the distinction between vulnerable and maintained abilities. Limitations and suggestions for future research are discussed.

Acknowledgements

I would first and foremost like to thank my thesis supervisor, Dr. Richard Fletcher, for his immense contribution, and for his warming words of praise throughout the course of this writing. I would also like to give a very special mention to Dr. Paul Wood of OPRA Group Ltd. for the immeasurable support and intellectual stimulation that he has kindly offered, and oft-forced upon me, over the course of the year. I am also earnestly thankful for the gesture of goodwill from OPRA Group Ltd, in particular Dr. Sarah Burke and Amy-Kate Snowise, for their support and also for allowing me access to the GRT2 New Zealand norm group data. Furthermore, I would like to proffer a big thanks to Psytech UK for access to the GRT2 assessment. Finally, I offer a big thanks to my mother for her unremitting words of encouragement and, in jest, to my flatmates for their oft-unconstructive time-management suggestions. Thank you.

Table of Contents

List of Tables

List of Figures

Introduction

> *"One of the most successful undertakings attributed to modern psychology is the measurement of mental abilities. Though rarely appreciated outside academe, the breakthrough in objectively gauging the nature an range of mental abilities is a pivotal development in the behavioural sciences"* (Lamb, 1994, p. 386).

Despite an extended history of debate and research, psychology is at pains to offer a universally accepted definition of intelligence. In fact, R.J. Sternberg noted that "viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it" (Gregory, 1998; Miller, Myers, Prinzi, & Mittenberg, 2009; Wang & Kaufman, 1993). It seems the notion of intelligence is intensely and intricately entangled within one's own method of inquiry, of which none appear to offer an interpretation that is more veridical than another. This thesis will address the subject under the pretext of the 'psychometric approach' to intelligence, specifically, with an age-related focus on Cattell's (1957) hierarchical model of intelligence.

Origins of the notion of psychometric intelligence and intelligence testing

The Ancient Greeks, specifically Aristotle and Plato, are well known for the earliest documented *conceptualisation* of intelligence (Jensen, 1998), however it is only in the past 100 years in which we have attempted to *measure* it (Eysenck & Fulker, 1979). While it could be argued that Sir Frances Galton was the first to measure 'intelligence' in some form or another, it was the French psychologist Alfred Binet and his associate Theophile Simon (Boake, 2010) whom created the first intelligence test in the form we know it today (Binet & Simon, 1916). Binet and Simon believed that Galton's simplistic conception of intelligence—as being a function of sensory acuity, physical features, and response times—did not offer adequate explanation for the extant complexity of intellectual ability. Binet's assessment was based on the notion that that intelligence surely *had* to be defined through some higher-order *mental* process (Kaufman & Lichtenberger, 2006; Murphy, 1968).

The structure of intelligence and the notion of 'g'

Giving structure to mental ability

It has been generally accepted that a hierarchical model of intelligence, generated through factor analysis, is the most suitable method for explicating the taxonomy of intelligence—theoretically, aesthetically, and statistically speaking (Jensen, 1998). Other models of intelligence, of which will not be discussed in this thesis, include the Spearman Model and the Thurstone Model (Thurstone, 1938).

In the hierarchal model of mental ability—a general example is shown in Figure 1—the *g* factor is said to be extracted out 'on-top' of the model through the correlation of the first-order group factors (F1, F2, F3) (Jensen, 1998). (For sake of demonstration these could be: Verbal, Numerical, and Abstract ability.) In the figure below, three first-order factors have been extracted from the nine items (V1-V9). A second order factor (*g*) is seen on the upper level of the model. This second-order factor has been extracted as a common source of variance between F1-F3. Variables V1-V9 are correlated with the second-order *g* only through their correlation with F1-F3 (Jensen, 1998).

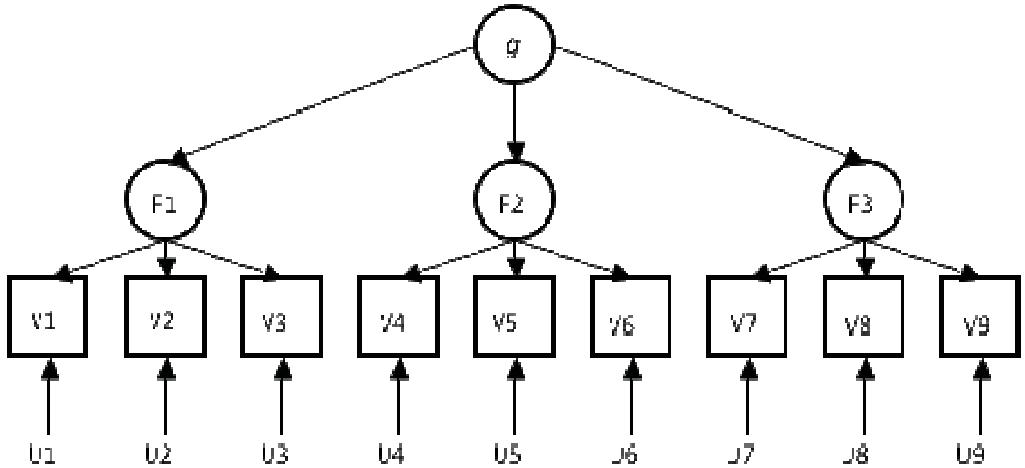

Figure 1. Example Hierarchical Model of *g.*

An American researcher, whom was not particularly enthused with the concept of a general factor, nor with orthogonal rotation, was Raymond Cattell. Cattell (1941) posited that there were two second-order factors that were derived from the factor analysis of first-order factors. Cattell proposed that when the traditional 'g' factor was rotated, not orthogonally, but obliquely (thereby allowing correlation), it split into two correlated factors. Cattell labeled these two factors Crystallised and Fluid intelligence (Cattell, 1941).

Cattell's theory of Crystallised and Fluid Intelligence

Cattell's theory of Crystallised and Fluid intelligence holds that cognitive ability can be organised at a general level into two core dimensions (Cattell & Horn, 1978; Horn & Donaldson, 1976; Horn & Cattell, 1982; Horn & Hofer, 1992; Horn & Cattell, 1967). The first, Fluid intelligence, is thought to be representative of the mental processes involved in reasoning on tasks necessitating such things as concept formation, abstraction, attainment, and the perception of relations (Cattell, 1963; Horn & Cattell, 1966a). Being said to exhibit itself more in tests that require adaption to new and novel situations (Cattell, 1963), it is thought that this type of intelligence is best measured on assessments that are 'culture fair' (Horn & Cattell, 1966a).

The other core ability of Cattel's two-factor model of intelligence is referred to as General Crystallised intelligence (abbreviated Gc). Unlike Gf, Gc is thought to be a product of acculturation, education, and experiential influences (Cattell, 1963; Horn & Cattell, 1966a). The Crystallised intelligence function infers the extent of information awareness and perception of subtlety in relational distinctions, seen in tasks necessitating recall or recognition of such awareness or relations (Cattell, 1963). Crystallised intelligence is also said to denote the ability to reason in the immediate environment in tasks requiring abstracting, concept formation and attainment, and perception (Horn & Cattell, 1966a). A comparison of the two, in more simplistic terms, views Fluid intelligence as reflecting the fundamental, innate reasoning capacity of the individual, whereas Crystallised intelligence is viewed as the knowledge base of an individual.

Investment theory

A particularly fascinating aspect of this theory is *Cattell's investment theory of intelligence*. Investment theory proposes that at the outset of child development there is a single 'relation-perceiving ability' associated with the maturation of cognition, that is: Fluid ability (Cattell, 1987; Valentin Kvist & Gustafsson). Cattell proposes that this Fluid ability is *invested* in the individual's growth and subsequent learning experiences. It is therefore posited that Gf moderates the rate of learning on tasks demanding insight and forethought, such as reading, arithmetic, and abstract reasoning (Cattell, 1987). Ergo, it is this investment of their Fluid ability, that results in acquiring knowledge and skills, i.e., acquiring Crystallised ability (Valentin Kvist & Gustafsson). The theory suggests that children with a higher initial level of Fluid intelligence have more cognitive power to 'invest' into skills and knowledge at the outset—usually those considered valued within a culture, e.g., in a Western culture this might be into scientific and technological subjects (Kline, 1998; Valentin Kvist & Gustafsson, 2008). It is for this reason that Crystallised and Fluid intelligence are significantly correlated within early development, but become less so as increasing amounts cultural experience, physical aging, and education affect intellectual development (Kline, 1998).

<u>The evidence for Crystallised and Fluid Intelligence</u>

According to Carroll (1993), the theory of Crystallised and Fluid intelligence offers "the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities" (p. 62). There is a variety of evidence that supports this claim, of which is manifested by two main forms of research: structural and developmental (Horn & Blankson, 2005). Structural research is seen in the study of covariation among assessments designed to measure forms of human intelligence, whereas developmental research looks at the ways in which intellectual capabilities progress over age.

Structural evidence

Structural evidence supporting the existence of Gf-Gc comes from factor-analytic data that demonstrate *real* individual differences. The structural evidence suggests that the described Gf-Gc components are genuine, in that they are reliably distinct from one another. The principle evidence comes from the seminal work of Carroll's *Survey of Factor Analytic Studies* (1993). Carroll critically reanalyzed over 400 data sets relating to cognitive ability scores. As an example of how great this achievement was, Jensen, a veritable authority within the discipline of psychometrics stated:

"Carroll's magnum opus thus distills and synthesizes the results of a century of factor analyses of mental tests. It is virtually the grand finale of the era of psychometric description and taxonomy of human cognitive abilities. It is unlikely that his monumental feat will ever be attempted again by anyone, or that it could be much improved upon. It will long be the key reference point and a solid foundation for the explanatory era of differential psychology that we now see burgeoning in genetics and brain sciences" (Jensen, 2004, p. 5).

Carroll's meticulous and venerated exposition into the structural representation of cognitive abilities elucidated 87 distinct elementary mental capacities, of which Carroll referred to as Stratum I, and which are now often considered the *primary mental abilities.* Factor-analytic evidence also elucidated eight or nine broader abilities above the primary mental abilities, which Carroll deemed 'Stratum II'. Crystallised and Fluid intelligence were represented, among others, on this broad second-order level (Carroll, 1993; Horn & Blankson, 2005). Table 1 offers an overview, with descriptions, of the relationships between the eight second-order (Stratum II) factors associated with each of the first-order (Stratum III) abilities.

Cattell (1987) offers further structural evidence for the support of Crystallised and Fluid ability. Results from the first *culture-fair intelligence* tests (Cattell, 1940)—of which are well supported as the practical expression of Fluid intelligence—point to a very different standard deviation of IQ, than do tests measuring *g* in its entirety. Cattell posited deviations up to 24 points, whereas, of course, typical *g-related* IQ assessments have standard deviations of 15, suggesting that the Fluid ability instrument was measuring something distinct from *g*.

Table 1. Carroll's (1993) second-order factors

Second-Order Factor	*Symbol*	*Definition*
Crystallised intelligence	*Gc*	*An ability measured in tests assessing breadth of knowledge regarding concepts, language, and other culturally-dominant information*
Fluid intelligence	*Gf*	*An ability measured in tests requiring reasoning, insofar as identifying relations, understanding inferences, and making inferences on equally over-learnt or novel content.*
Working memory	*SAR*	*Also referred to as short-term apprehension and retrieval, this is a measure of a variety of tasks that require the awareness of elements in the immediate environment for a short period of time.*
Fluency of retrieval from long-term storage	*TSR*	*Also referred to as long-term memory, this is the ability to consolidate, store, and recall information stored for longer periods of time.*
Processing speed	*Gs*	*An ability of which is measured in timed tasks requiring the participant to correctly respond as quickly as they can to a simple task.*

Visual processing	*Gv*	*An ability measured in tasks requiring the recognition of the way shapes are rotated in three-dimensions.*
Auditory processing	*Ga*	*An ability measured in tasks requiring the acute perception of auditory patterns. More specifically, the awareness of order, rhythm, and elements of groups of sounds.*
Quantitative knowledge	*Gq*	*An ability measured on tasks requiring the understanding and application of numerical concepts relating to mathematics.*

Note: some authors, such as Flanagan and McGrew (1998) also propose an additional factor of Reading-Writing (Grw), of which concerns the ability to comprehend written text and communicate through written text. Gc is sometimes referred to as Acculturation-knowledge (Horn & Blankson, 2005).

Age-related developmental evidence

Although the structural evidence indicates the independence of Gf-Gc through covariation and factor-analysis, age-related developmental evidence highlights the independence of Gf and Gc through their unique interaction with age. The reasoning of this argument follows simple logic: if two apparently distinct 'factors' change at the same rate, it suggests they *could* be one in the same, whereas if they change at a *different* rate, this indicates with almost certainty that the constructs are distinct from one another. Many studies have demonstrated the phenomenon that both Crystallised and Fluid intelligence have distinctly different interactions with age (e.g., Cattell, 1987; Horn & Cattell, 1967; Kaufman & Lichtenberger, 2006). Fluid ability has been shown to peak at around the age of 20 and decline steadily therein, whereas Crystallised ability seems to rise steadily from birth and peak at around 65 years (Berowitz & Green, 1963; Bugg, Zook, Delosh, Davalos, & Davis, 2006; Cattell, 1987; Horn & Cattell, 1967; Kaufman & Horn, 1996; Kaufman, Reynolds, & Mclean, 1989; Kaufman, 2001; Mcgrew, 2005; Ryan, Sattler, & Lopez, 2000; Salthouse, 2004).

Other developmental evidence comes from brain damage studies, whereby brain damage affects performance on the two abilities differentially. For example, damage to the Broca area of the brain can produce aphasia—partial or total loss in the ability to articulate ideas or understand spoken or written language—but no apparent loss in non-verbal related performance (Adams & Victor, 1993). Conversely, any damage to cortical areas invariably results in at least *some* damage to Fluid ability performance (Reitan, 1955, and Lashley, 1963, as cited in Cattell, 1987).

In summary, the structural evidence for existence of these factors as real and distinct, comes from distinct construct validities and solid developmental differentials, in which they both have differing relationships with other variables, such as neurological functioning, behavioural genetics, occupational performance, school attainment, and age (Horn & Blankson, 2005).

Criticism of Crystallised and Fluid Intelligence

One author (Guilford, 1980) criticised Cattell's hypothesized Fluid and Crystallised intelligences as not being able to stand up as the two broad abilities Cattell thought them to be. Guilford posited that these two "factors" were at the very most, representative of two different types of test. He claimed that the conditions of which Horn-Cattell (1966) designed to provide the distinction between Gf and Gc were defective because they were especially favorable for illustrating these 'second-order factors'. Guilford was particularly concerned about Horn and Cattell's (1966) sample that had a controlled variance on chronological age (14 to 61 years) but was heterogeneous in all other regards, such that he posited the basic rules of experimental control were violated. Guilford therefore speculated that the given Gf and Gc factors were spurious in nature. Horn and Cattell (1982) rebuked these charges sharply in two ways, in which demonstrated that the sample construction methods were appropriate for the purpose of the initial study, and that if the sample construction was adapted to fit Guilford's need for having experimental controls, that the results would remain congruent with the initial results.

No similar claims similar to Guilford's have emerged since Horn and Cattell's reply. In fact, that the Gc and Gf framework has been subsequently supported by the likes of Carroll (1993) has veritably strengthened support for the model, and has thus seemed to quell further reproach.

Cattell-Horn-Carroll theory of human cognitive ability

It could be said that the two most commonly accepted hierarchical models of cognitive ability are Horn and Cattell's theory of Crystallised and Fluid intelligence (see e.g., Horn and Cattell, 1966) and Carroll's (1993) three stratum theory, developed from his seminal meta-analyses study). There are significantly noteworthy similarities between the Gf-Gc theory and Carroll's three-stratum theory, so much so a single theory was developed to minimise redundancy in the discipline and encompass them both, viz., the *Cattell-Horn-Carroll* (CHC) *theory of intelligence* (Cole & Randall, 2003; Mcgrew, 1997); a model which is believed to act as the tent that houses the two most important psychometric theoretical models of human cognitive ability (Daniel, 1997; Mcgrew, 2009; Sternberg & Kaufman, 1998). McGrew (1997) proposed this amalgamation with the justification that there was a need for "a single broad and narrow ability taxonomy by which to classify the narrow abilities measured by the tests in the major individually administered intelligence batteries" (McGrew, 2009, p. 3). The broad CHC abilities are posited to be: Gf, Gc, auditory processing (Ga), visual processing (Gv), short-term memory (Gsm), long-term memory (Glr), processing speed (Gs), quantitative knowledge (Gq), and reading and writing (Glr). Support for the existence of these nine variables has been consistently demonstrated through small and large sample structural validation studies (Mcgrew, 2005; Stankov, 2000). It is now thought that the McGrew CHC model represents the most accurate model of factors underlying the most commonly measured facets of intelligence (Kaufman, Johnson, & Xin Liu, 2008; Mcgrew, 1997; Mcgrew, 2009).

A pure Gf and Gc?

Some authors have suggested that many studies, such as the present, that have taken a purely Catellian view of intelligence—that is measuring only Gf and Gc—are likely to be inadvertently contaminated by other broad abilities of the CHC model. An example is a seemingly pure timed measure of Gf, such as a visual-spatial manipulation item that requires the testee to make the inference as to which stimuli logically follows in a series. On closer inspection, this measure of Gf may also be reflective of ones' processing speed (Gs), visual processing ability (Gv), and short-term memory (Gsm); an ostensibly plausible proposition. A timed test requires one to process information with urgency (identify next pattern in series as quickly as possible), process *visual* information (perform cognitive visual-spatial manipulations with the items presented through visual medium), and use short-term storage and retrieval memory processes (perceive and store the hypothesized stimuli-relationships for testing). The degree to which the Gf and Gc factors in the current investigation will be impacted by these extraneous variables is largely unknown, but will be considered in the analyses.

<u>The Measurement of Crystallised and Fluid Intelligence</u>

Measurement of Fluid Intelligence

Cattell and Horn (1978) espouse several key principles for designing tests of Fluid intelligence: (1) The assessments should be reflective of the 'eduction of relations and correlates' (Cattell & Horn, 1978; Spearman, 1927), and be of an increasing hierarchical difficulty (under the classical testing model); (2) although being based—necessarily on the discrimination of visual-spatial figures, the tests should base their fundaments of difficulty on higher-order non-verbal relations, such as part-whole relations, super-ordinate relations, developmental relations, and degrees of similarity; and (3) the tests should be free from any fundaments in which differential ability to understand the items stems from knowledge or cultural symbolism of some form. Ergo, measures of Fluid ability are commonly measured through tests that have virtually no cultural or education-based content. The most common form of Fluid intelligence test are perceptual and figural tests, such as Raven's progressive matrices (Jensen, 1998) (see Figure 2). This type of perceptual or figural test usually contains a series of shapes, requiring the testee has to identify which shape would follow logically from the others. Jensen (1980) gives other Fluid ability item examples:

- Figure classification. Similarities of a series of figures must be identified.
- Embedded figures. Certain shapes in a figure containing many other shapes set inside one another must be identified.
- Reoriented shapes. Successful mental rotation of geographic shapes.
- Figure identification. Pictures of blocks are determined as: same or different.

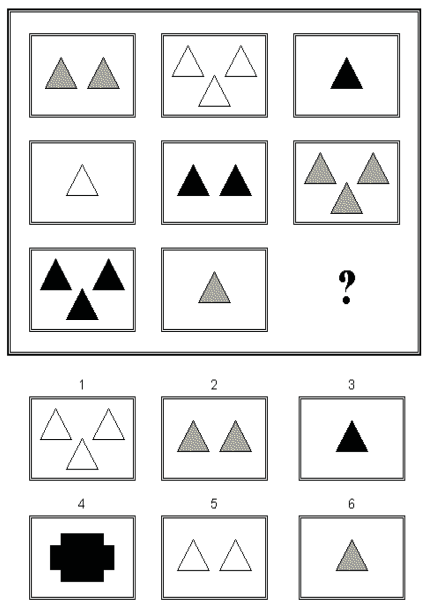

Figure 2. Example matrix relations item. Reprinted from "Raven's Matrices mock-up" by S. Bell (2003). Reproduced from http://www.stuartbell.co.uk/ravens.html, accessed 23rd August, 2011. Reprinted with permission.

Measurement of Crystallised Intelligence

Unlike Fluid intelligence tests, tests of Crystallised intelligence make knowledge demands upon the testee (Cattell, 1987; Jensen, 1998), such that the individual level of educational opportunity and cultural knowledge acquisition is being assessed.

The three most prominent item types which characterize measures of Western Crystallised intelligence are verbal ability, general information, and reading comprehension (Kline, 1998). There is a clear case for why these three types of items characterize Crystallised intelligence to such an extent. Recall how Cattell defined Crystallised intelligence as the investment of Fluid intelligence into the skills of which are valued in a culture (Cattell & Horn, 1978). The intellectual skills valued in our Western Culture are essentially verbal, and thus the development of verbal skills is likely to be valued (Kline, 1998). The fact that much of the performance requirements in the Western education system, and increasingly the Eastern system, are verbal by nature (e.g., written exam, spoken presentation, written assignments) supports this notion. Thus it could be argued that very little of Westernized schooling assesses and reinforces the learning of purely non-verbal performance; even mathematics often requires the student to explicate their reasoning verbally.

The typical verbal ability test may be comprised of these sorts of items: synonyms, vocabulary, verbal analogies, verbal classifications, and proverbs. For instance, an example of a verbal classification type test item of Crystallised ability would be:

‘Which of the following is the odd one out?’

a) Apple b) Orange c) Tomato d) Mandarin e) Carrot f) Avocado

In this example the knowledge of the difference between vegetables and fruit is required, specifically being able to distinguish the vegetable (i.e., carrot) from the fruit. Jensen (1980) and Cattell (1987) cite further examples of the types of verbal tests that assess Crystallised ability:

- Verbal classification tests: see above example.
- Verbal similarities: e.g., '*hungry* means the same as…?'
- Similar opposites: e.g., '*dull* means the opposite of…?'
- Syllogisms: e.g., 'if all apples are red, and all fruit is green, then are apples fruit?'
- Scrambled sentences: 'unscramble the following proverb: eht relay bdir tges eht wrom'.
- Sentence completion, e.g., 'in his usual ____ manner, he had insured himself against this type of loss'.

Relationship between Fluid and Crystallised ability

It is often the case that a clear-cut distinction between whether a test item loads on Fluid or Crystallised intelligence is not completely evident. For instance, analogies can have a tendency to load onto both Crystallised and Fluid intelligence. Consider the following two items:

1. **Height** is to **tall**, as **Temperature** is to:

a) Hot b) Inches c) Weight d) Size

2. **Green** is to **Envy** as **Red** is to:

a) Sky b) Anger c) Yellow d) Envy

Because they are both examples of *verbal* analogy, one might expect them both to load on Crystallised intelligence exclusively. However, upon closer examination we could predict the first example might be likely to significantly load onto both Crystallised and Fluid intelligence, whereas the second example would be more likely to load more significantly onto Crystallised rather than Fluid. While the first contains basic verbal information, for which we must have specific knowledge in order to solve (i.e., commonly known measurement terms), it also requires a particularly Fluid-like ability to perceive exactly how, and on what level, the words relate to one another. One must first identify that both 'height' and 'temperature' are ratios of physical quantities and second that 'tall' is an adjective that is used to describe height. The inference must then be made that the correct response in the item response options would be an adjective that can be used to describe 'temperature' (i.e., option 'a', 'hot'). The second example does not require this degree of eduction of relations, as it merely requires the testee to realize the learnt association between the words 'green' and 'envy' and the idiom 'green with envy', a phrase used to describe a certain human condition within the English language. Once this association has been made the testee would then have to 'scan' their knowledge banks to look for a colour idiom that described a human condition and is associated with the word 'red'.

Notice that in this cross-loading example we are concerned about a supposedly pure Crystallised item loading onto both Crystallised ability and Fluid ability factors. It is well demonstrated that Fluid ability has a tendency to enter into the verbal ability primary and sometimes the numerical and reasoning primaries (Cattell, 1987). Thus it appears some degree of relation eduction and adaptability to new situations—such as characterized by Fluid intelligence—is required even for situations that appear only to use acquired knowledge (Cattell, 1987; Horn & Cattell, 1967). It is for cases such as these that we must be prudent in not readily assuming that a verbal item will load exclusively on to Crystallised intelligence.

All of the above supports the idea of a notable correlation between Crystallised and Fluid intelligence; indeed research has deemed this assertion appropriate: Gf and Gc have been demonstrated to positively correlate significantly at all ages. Cattell (1987) suggests this correlation to be around the 0.4 to 0.5 vicinity.

Population differences in cognitive ability

Crystallised and Fluid intelligence are strictly phenotypic, that is Gf and Gc can only ever be a physical manifestation. Gf and Gc alone are not sufficient in providing support for any inference pertaining to the cause of individual or group differences (Jensen, 1998). Indeed it is no secret that both have shown to be significantly correlated, to differing degrees, with economic, social, and educational factors (Jensen & Rushton, 2005; Neisser et al., 1996). It is for this reason that we have seen prolonged interest in the cognitive ability differences between groups that differ on these economic, social, and educational characteristics. Three demographical differences of which have been parsed with much vigor are ethnicity, gender, and age.

Age-related differences in cognitive ability

The phenomenon of cognitive aging has been noted for as long as we have been aware of *physical* aging, but it is still not well understood (Salthouse, 2004). A number of age-related changes in cognitive ability have been studied, including: memory, processing speed, and reasoning. The changes relative to each have been well documented—and generally well supported—over the lifespan of intelligence research (Kaufman, 1990; Ryan et al., 2000; Schaie, 1994). The study of age-related differences in mental ability dates back over a century ago to 1903 (Kirkpatrick, 1903) and has since been extensively investigated within psychological literature (e.g., Cattell, 1987; Horn, 1967; Jones & Conrad, 1933; Kaufman, 1990; Kaufman, 2001; Kaufman et al., 2008; Lorge, 1936; Miles & Miles, 1932; Miller et al., 2009; Schaie, 1958). While the analysis of thesis will be based on the change in Crystallised and Fluid intelligence, it is in the interest of a more in-depth understanding that the cognitive decline research surrounding the Wechsler Adult Intelligence Scale (WAIS) and Primary Mental Abilities (PMA) series of ability tests be detailed first. As will be covered, strong parallels can be made between the WAIS and PMA measures and measures of Crystallised and Fluid intelligence.

The WAIS body of cognitive decline research

The WAIS series of assessments are considered the most up-to-date, well-validated, and most accurate assessments in the field of cognitive decline (Kaufman, 1990). In fact, Kaufman (1990) claims that the studies conducted using the WAIS (e.g., Kaufman & Lichtenberger, 1999; Kaufman, 2001; Miller et al., 2009; Ryan et al., 2000; Sattler, 1982) offer the most valuable and valid research findings for the interpretation of age-related cognitive decline. Out of all cognitive ability assessments, the WAIS also has the greatest wealth of supporting research relevant to this cause.

A cross-sectional WAIS approach to age-related cognitive decline

The largest body of WAIS research into cognitive decline is cross-sectional. Before these findings are discussed however, it is important to note the major limitation of age-related cross-sectional research in general, known as the cohort effect. Groups of increasing chronological age unavoidably differ on further variables that may act to confound the true between-group age-related variability. It is a fact that a child growing up in the 1920's would have necessarily undergone a significantly different upbringing than would a child who was born in the 1990's. An upbringing in the 1990s, as opposed to the 1920s would be associated with greater amount of education, informational opportunity, greater nutritional standards, and a culture which arguably treats both male and females with equal right in obtaining such (Kaufman & Lichtenberger, 2006). It is therefore inevitable that results from a cross-sectional study will be influenced by a function of chronological age *and* these 'cohort' differences exemplified by the period over which the individuals lived.

Contemporary researchers are now well aware of the impact of cohort dynamics, yet early researchers were not. Kaufman & Lichtenberger (2006) posit that generational effects were mostly ignored by researchers through the 1950's and though most of the 1960's. Even Wechsler (Wechsler, 1939, 1958) appeared to over-infer a rapid degradation of intelligence as he perfunctorily accepted that the decline in mean-scores across his standardization samples was caused solely by cognitive decline.

The data appeared to suggest that the older participants had a consistently lower IQ than younger participants (Kaufman & Lichtenberger, 2006).

Further insight is obtained when we look at the sample demographics of these early WAIS studies. In particular, Kaufman & Lichtenberger (2006) draw attention to 'years of schooling' variable for the standardization samples that Wechsler used in the development of the WAIS, WAIS-R, and WAIS-III (1953, 1978, 1995 respectively). Kaufman and Lichtenberger's (2006) analyses looked at the percentages of subjects in each standardization sample whom had undergone 0-8 years of schooling, compared with the proportion whom have had 13+ years of schooling. The results demonstrate that the data was initially misinterpreted, as Kaufman & Lichtenberger's analysis demonstrated significant educational differences within and between each of the three WAIS samples. To underscore but one example: 49 percent of subjects aged between 45-54 in the recent 1995 standardization held 13+ years of schooling, contrasted with a paltry 14 percent of subjects of the same age group in the 1953 sample. It is now clear to researchers in this discipline that with each passing decade, an ever-increasing proportion of individuals are staying longer at high school, and now more than ever, attending university (Kaufman & Lichtenberger, 2006; Ministry-of-Education, 2010). It is this fact that seems to be able to explain why younger age groups will tend to be relatively more "intelligent" than older adults. Perhaps the phenomenon of cognitive decline is illusory; perhaps the decline is only a function of level of education?

That 'cognitive decline' was simply a function of educational opportunity was entertained by Birren and Morrison (1961). Birren and Morrison addressed the possible education confound by statistically controlling for years of formal education. All of the 11 intelligence dimensions in their study initially correlated negatively with age, with the Performance dimensions (analogous to Cattell's Fluid intelligence; Matarazzo, 1972) demonstrating the strongest negative correlations with age. Through statistically partialling out education level, six out of eight of the Verbal dimensions (analogous to Cattell's Crystallised intelligence; Matarazzo, 1972) produced *positive* correlations, suggesting an *increase* in test scores with age. The two Performance dimensions retained the negative correlations with age, following the partialling of education level, suggesting that IQ, with respect to WAIS Performance dimensions and Cattell's Fluid ability dimension, decreases with age.

Kaufman, Reynolds, and McLean (1989) investigated the same research question through the analysis of the WAIS-R standardization data, for ages 20 to 74 years, and with the sample being carefully compiled with respect to gender, race, geographic location, educational attainment, and occupation. Using an ANOVA type statistical analysis, the authors controlled for educational attainment by holding the effect of education constant through a target age range between 25-34 years. They also conducted a second analysis using a multiple regression approach to elucidate the relative contributions of age and education. This was to ascertain whether age added anything significant to the prediction of cognitive ability over and above the contribution of education. After controlling for education, the previously apparent decline in Verbal ability all but disappeared. Nevertheless, the decline in Performance scores remained significant. After equating for education, Verbal ability peaked for the 55-64 year old group, and declined insignificantly thereafter;

whereas Performance ability declined significantly (at a 0.01 level of significance), in a linear fashion, from a mean IQ of 103 for the 20-24 year group right through to a mean IQ of approximately 76 for the 70-74 year group. The results of Kaufman et al. (1989) ANOVA analysis are entirely paralleled by the earlier Birren and Morrison (1961) data. This finding provides even more support for Crystallised intelligence showing a general increase and Fluid ability showing a general decrease, across age.

Furthermore, Kaufman et al.'s (1989) multiple regression analysis offered additional support for the education hypothesis. When added initially to the model, education accounted for nearly half the variability of Verbal ability, and nearly a third of Performance. This result was expected; instead, what was of more interest was whether chronological age would add significantly, and substantially, to the prediction model when added as a second predictor. Adding age as a predictor resulted in a 13% increase in explained variance for Performance IQ, but only a 0.3% increase for Verbal IQ. Thus, the notion that lower scores on Verbal subtests for older age groups do not persist when education is controlled for is supported. Declines on scores for Performance subtests, however, persist even after controlling for the potentially contaminating influence of educational attainment.

The above research conforms closely to Horn and Cattell's (e.g., 1966, 1967) predictions of Gf and Gc's age-related changes. As suggested by Cattell, Crystallised ability can be thought of as a "maintained" ability of which increases through the 60s before beginning a decline in the 70s (Baltes, 1997; Horn, 1989); whereas Fluid ability, can be thought to exist independently of education and acculturation, and can be thought of as a "vulnerable" ability, which peaks in the early 20s declines linearly there onwards.

To further support this line of cognitive decline research using the WAIS series of assessments, Kaufman (2001) investigated the age-related changes in cognitive ability as measured by the WAIS-III, the improved version of the WAIS-R. Through using similar analyses to Kaufman et al. (1989), the results once again demonstrated further support for the notion of both "maintained" and "vulnerable" intelligences. That is, across age, Verbal ability was maintained and Performance was vulnerable (Kaufman, 2001).

Summary of cross-sectional investigations into changes in intelligence across age

Most of the cross-sectional research into cognitive decline has been based on the Wechsler Adult Intelligence Scales (WAIS) series of assessments, with particular emphasis on the control of education and cohort differences. Very occasionally results have suggested a maintained level of Fluid ability with advancing age (e.g., Green, 1969), however a prevalence of data suggests a somewhat steady and significant decline of Fluid ability with advancing age. Although we might blindly accept this suggestion as fact, we must be cognizant of the inherent limitations of cross-sectional investigations.

A longitudinal WAIS approach to the question of age-related cognitive decline

Longitudinal investigations eradicate the problems associated with cross-sectional approaches by following the development related changes of the *same* individual across time. By following the same individual across time we are effectively holding cohort variance constant, as each individual is essentially his, or her, own control. The very few longitudinal investigations of ability have thus far, interestingly, shown little age-related decline in intellectual ability (Crystallised or Fluid) (e.g., Berowitz & Green, 1963; Eisdorfer & Wilkie, 1973) even when controlling for education.

Acknowledging the major weaknesses of cross-sectional studies, one would hope that by following individuals' level of Gf and Gc over time would allow us to either support or refute the notion of intellectual decline, but research to date has yet to cogently answer this question (Kaufman & Lichtenberger, 2002). It is the case that longitudinal investigations into age-related cognitive ability are entrenched with two of their weaknesses. The first is that longitudinal studies in this field tend to call for the same instrument to make repetitive measures across time, thereby introducing unwanted error. This is a problem known as *progressive error* (Kausler, 1982). This error is normally associated with the nature of the items and the practice effects that are associated with them. For instance, individuals retested on the WAIS-R following a month are likely to gain a mere 2 or 3 points on the Verbal subtest, whereas on the Performance Scale they are likely to gain around 8 or 9 points (Kaufman, 1994b). Indeed, this overwhelming increase is said to persist for *at least* 4 months following testing sessions (Catron & Thompson, 1979). Therefore, the longer the temporal space between test and re-test, the less progressive error would be an issue. Using Catron and Thompson's data, it would follow that the progressive error for longer periods (e.g., 5-10 years) would be negligible.

The second major issue with longitudinal research is the selective attrition of participants. It has been demonstrated that those who do not volunteer for the study and those who do not show up in the subsequent retesting tend to be of a lower cognitive ability in the first place (Horn & Donaldson, 1976). Eisdorfer and Wilkie (1973) demonstrated a particularly pertinent example of this through examining dropouts of the Duke University studies. The lowest IQ group suffered a loss of 72 percent of the participants over the course of the study, the middle IQ group 51.4 percent, and the high IQ group only 36.8 percent. In fact, a further study by Sieger and Botwinick (1979) demonstrated an almost linear relationship between IQ at the initial assessment and the length with which the participants willingly returned for retest sessions. Undeniably, the combination of these two issues offers a lucid testament as to why longitudinal studies must be interpreted with prudence. With this in mind, two seminal longitudinal age-related decline studies are reviewed.

In 1950, Owens delivered the Army Alpha assessment to 127 men (aged 50) whom had completed the test thirty years prior, in 1919. Ninety-six of these individuals were then retested in 1961, alongside a new sample of 19 year-old Iowa freshmen students, as to estimate the amount of cohort-related cognitive change between 1919 and 1961. When the cohort-unique differences were controlled for, Performance scores were shown to significantly *decrease* across age, as has been repeatedly illustrated in similar cross-sectional research. Verbal scores conversely, demonstrated steady gains between ages 19 and 50, followed by a slight decline between 50 and 61.

Schaie (1983b) conducted perhaps the most sophisticated age-related cognitive studies to date. Using the group administered Primary Mental Abilities assessment, Schaie conducted four cross-sectional studies in parallel with longitudinal studies.

This rigorous combination of cross-sectional and longitudinal studies allowed Schaie to control for cohort-unique differences and time-of-measurement variation, thereby supposedly permitting Schaie an attempt at identifying real intelligence differences associated with aging (Kaufman & Lichtenberger, 2002). Unlike Owen's Performance data, Schaie's Performance data did not conform to the cognitive aging pattern seen in previous cross-sectional research, nor that seen in Owen's (1966) research. Schaie's data suggested that Performance ability was relatively stable right throughout middle age and that it begun its decline after about 60 years. This is certainly at a far later age than suggested by previous research (e.g., Horn & Cattell, 1967; Kaufman & Horn, 1996; Kaufman & Lichtenberger, 2006; Kaufman, 2001; Kaufman et al., 2008; Owens, 1966).

To further investigate this finding, Kaufman (1990) used a cohort substitution method on the WAIS, WAIS-R, and WAIS-III normative samples to ascertain whether longitudinal data would replicate the sharp early 20s decline in Performance ability seen in similar cross-sectional data and in Owen's longitudinal data. The results were unequivocal, on both studies Verbal ability was supported as a maintained ability (Cattell, 1987): a peak in ability emerged in the 45-54 year age group and therein showed only a small decline. Interestingly, Performance IQ once again appeared as a vulnerable ability (Cattell, 1987): peak levels emerged in the 20-24 year group and declined at a rate of about 5 points per decade, which closely mirrors that of Horn et al.'s (1985) estimations of Fluid intelligence decline.

But what can reconcile these differences in outcome between Schaie's PMA data and Kaufman's Wechsler data? The remarkable post 20-24 year Performance decline seen in the WAIS data is not observed in Schaie's PMA data. Why? Kaufman and Lichtenberger (2006) believe that if Schaie had used the WAIS instead of the PMA, Performance would have been maintained and that the difference may relate to the sensitivity and complexity that characterises Wechsler's Performance measures. These authors speculate that it is this sensitivity and complexity, which is said to be lacking in the PMA, that makes the WAIS tests highly responsive to inflictions of age-related ability decline (Kaufman & Lichtenberger, 2006). If one thing is clear, it is that there is strong support for notion of a vulnerable Fluid ability, and a maintained Crystallised ability. This has been supported from both the perspectives of cross-sectional and longitudinal research.

A short note on the Flynn Effect

In the interest of a well-rounded discussion on cognitive aging patterns, the Flynn Effect must be mentioned. The Flynn Effect (Flynn, 1987; Hernstein & Murray, 1994) is a term given to substantial and sustained increases in intelligence scores across age cohorts. Richard Lynn first noticed the effect upon discovering the trend over subsequent intelligence assessment re-standardisations. Lynn(1982) noticed that following these re-standardisations, the sample score means (particularly for non-verbal) for the newer standardised data had to be adjusted downwards such that the mean IQ remained at 100. This seemed to suggest that the average population intelligence was increasing over cohort age. Indeed, these increases appeared continuous and approximately linear from when the phenomenon was first noticed up till the present (Flynn, 1987). Of further interest, these increases have been well documented across many countries. However, recent research seems to suggest that the Flynn Effect may have ended for many developed nations (Lynn & Vanhanen, 2006)

Moreover, the gains of the Flynn Effect seem particularly strong for those in the lower end of the intelligence distribution; with one study illustrating a preponderance of low-end scorers showing substantial increases, whereas the higher-end scorers remained relatively unchanged (Teasdale, 1989). Another study suggested that the increases appeared to have a moderate negative correlation with intelligence and that the increases could be explained by a decreased variance in the lower end of the distribution (Colum, Lluis-Font, & Andrés-Pueyo, 2005).

While the acknowledged findings of the Flynn Effect will be taken into account in the present investigation, it is likely that any potential Flynn Effect seen in the results will manifest through only slightly attenuated decreases in Fluid intelligence across age, and only slightly increased gains in Crystallised intelligence across age. It must be emphasised however, that New Zealand may well be one of the developed nations for which the Flynn Effect has ended. Indeed, the gains reported by Flynn (1987) detailed New Zealand as having one of the lowest IQ gain scores (0.242).

The present investigation

In the present study, cognitive decline is investigated from the vantage point of Cattell's (1941) initial distinction between Crystallised and Fluid intelligence and subsequent adaptations of this theory (e.g., Cattell, 1963; Cattell & Horn, 1978; Horn, 1989; Horn & Cattell, 1966a, 1967; (Horn & Hofer, 1992). Parallel to the findings of Performance and Verbal ability, Crystallised ability has consistently been shown to be maintained over one's lifetime, showing only small declines in later life; whereas Fluid ability has been repetitively demonstrated to decline steadily and often radically, beginning at young adulthood (20-24), through to old age (Birren & Morrison, 1961; Horn & Hofer, 1992; Horn & Cattell, 1967; Kaufman & Horn, 1996; Kaufman, 2001; Wang & Kaufman, 1993).

This current study therefore aims to replicate these findings in a New Zealand sample, using a cross-sectional methodology. This will be investigated by using the Graduate Reasoning Test (GRT2) New Zealand norm group data.

Method

The first aim of this research was to validate the measure with which the subsequent research was based on. The second aim of this research was to identify the form of the relationship that Crystallised and Fluid intelligence has with age in a New Zealand sample. This was investigated with a cross-sectional approach on an archived cognitive ability dataset.

Measure

The General Reasoning Test (GRT2)—developed in the United Kingdom by Psytech International Limited—was used as the cognitive ability measure. The three GRT2 sub-tests are: Verbal, Numerical and Abstract. Subtest scores are computed by summing the number of correct responses. Of the three sub-tests, only data from the Verbal and Abstract scales were used in the final analyses. This is because, of the three sub-tests, Verbal and Abstract most clearly separate into Crystallised and Fluid ability (Ackerman, 1992; Carroll, 1993).

Sub-test characteristics: Verbal and Abstract

Verbal (Gc). The Verbal score was based upon timed performance on 35 items. For each item the respondent is presented with a short question of approximately 8-10 words in length, with six possible response options, only one of which is correct. Typical of Crystallised ability assessments, this section assesses the candidate's ability to correctly understand the relationship between two or more words. For example, 'bright means the opposite of…?', 'Which of the following words is the odd one out?', or 'Large is to Size as Quiet is to…?'. This Verbal measure is taken as representing Cattell's Crystallised intelligence.

Abstract (Gf). The Abstract score was based upon timed performance on 25 items. The respondent is either presented with: a) a question such as 'Which of the following is the odd one out?' followed by six visual-spatial figures labeled 1 through 6; or b) a series of 4 visual-spatial figure, and the question, 'Which one comes next?'; or c) a question such as, '*This* figure is to *that* figure, as *this* figure is to *what* figure?' Once again, the respondent has 6 response options to choose from, all of which are visual-spatial figures. This Abstract measure is taken as representing Cattell's Fluid intelligence.

GRT2 administration

When sitting this assessment, participants were given the option of completing either on the computer or paper and pencil. Approximately 99% of individuals opted to use the computerised version. In terms of the standardisation procedures, Psytech ensure that: a standard instruction set, a standard respondent input process, a standard environment, and a standard time-restriction for each sub-test, are used. The standard time restrictions of 8 minutes for the Verbal section, and 10 minutes for the Abstract section were used.

<u>Data</u>

The data was collected from a New Zealand based GRT2 archive that had been compiled for the purpose of creating NZ norm groups; such that each of the respondents has at some time completed the GRT2 assessment as part of an employment selection process. Being targeted at the 'general' employee level, the GRT2 sample group can be considered as representative of the average NZ job applicant. The initial dataset was comprised of just less than 10,000 New Zealand respondents. The final dataset was comprised of 5097 respondents between the age of 16 and 64 (*M*=31.46, *SD*=9.26), of which, 2306 recorded themselves as male and 2791 as female.

All respondents whom completed the assessment prior to 2001 were removed because of a substantial amount of missing respondent age information. Respondents were also removed whom were outside of the age range of 16 to 64 years because there were too few respondents in these age ranges to be statistically useful.

Through missing-data analysis (results shown in Table 15 and Table 16 in Appendix B) it was determined that there was an increase in missing item data towards the end items of each sub-test. Although no statistically defined standard cut-off point was used, it was decided that the last 5-items on each sub-test were to be discounted from the subsequent analysis. Larger amounts of missing data for items towards the end of each scale can be put down to the fact that the GRT2 is a time-restricted ability assessment in which respondents are under pressure to answer as many items, as accurately as possible. The Verbal section allows approximately 17.15 seconds per item, and the Abstract section approximately 19.2 seconds per item. Furthermore, as suggested by Table 15 and Table 16 in Appendix B, some items appear to have slightly more missing data than others (e.g., AQ1 has 3.77% missing data). This is likely to be a product of the interaction between item-difficulty and suitability of response alternatives. Analysis of the distractor and item-difficultly interaction for each item is out of the scope of this thesis, and will therefore not be investigated further.

Age group stratification

Data was classified into 5 age groups. It was decided that the upper age group would be cut off at 58 years, because this was the last age group in which there was substantial amount of data. Older age groups contained only 22 cases in total, and could therefore not contribute to the analysis. The secondary reason for this was to maintain suitably-sized age groups with roughly equal intervals; incorporating the 59-64 age range in the upper group would have added an insignificant *n* to the analysis to the total sample size, all the while undesirably removing a degree of specificity. The five groups varied in size from *n*=195 for the 51-58 year age group, through to *n*=1934 for the 25-33 year age group, all of sufficient size for the subsequent analysis. While it was initially attempted to maintain identical age groupings to those of the WAIS standardisation data—as this is what the majority of age-decline research is based on—this could not be accomplished due to the limited amount of respondents of 60+ years in the current dataset. Age groupings were otherwise chosen as to best represent New Zealand population age-demographics, according to recent Census data, all the while keeping the number of groups to a parsimonious minimum for statistical purposes.

Table 2. Respondent frequency per age group

	Age-Group (Years)	Frequency	Valid Percent	Cumulative Percent
Valid	16-24	1376	27.1	27.1
	25-33	1934	38.1	65.2
	34-42	1083	21.3	86.6
	43-50	487	9.6	96.2
	51-58	195	3.8	100
	Total	5075	100	
Removed	59-64	22		
Total		5097		

Data Analyses

Internal consistency: reliability analysis

The current investigation investigated Cronbach's alpha for both sub-tests. An item-level reliability analysis was conducted in parallel to this, in order to identify and remove items that were decreasing measurement consistency. The final Cronbach's alpha reliability was also estimated.

Descriptive analysis

A descriptive analysis was conducted on the dataset. Individual items for both sub-tests were reviewed initially, and then the means were calculated for each sub-test across the five age groups.

Item-difficulty analysis

In order to more fully understand the GRT2, the selected items were analysed in terms of their difficulty level. Ideally the assessment would contain a large variance in item difficulty: between chance (for 6 options: 0.167) and approximately 0.9. This ensures a greater degree of ability discrimination to cover the ability scale: easy to hard.

One-factor versus two-factor model

The dataset was analysed to determine whether a one-factor or two-factor model would fit the data better. The one factor model is in line with general factor of intelligence theory, or *g* (Carroll, 1993). The two-factor model tested was representative of Crystallised (Gc) and Fluid (Gf) intelligence. Model fit was evaluated through using multiple goodness-of-fit measures deemed appropriate by Bentler (1990) (i.e., Comparative Fit Index (CFI) (Bentler, 1990), Normed Fit Index (NFI) (Bentler & Bonett, 1980), and Root Mean Square Error Approximation (RMSEA)). While there is no absolutely pre-defined levels of fit, Byrne (1994) suggests CFI be above 0.93, NFI be above 0.90, and Browne and Cudeck (1993) recommend that RMSEA be below 0.08. Although these cut-offs are recommended, they will be interpreted in the context of the current analysis (Kline, 2000).

Assessing age-related ability differences and variance

Assessing the relationship between age and Crystallised and Fluid intelligence was the primary aim of this thesis. In an attempt to measure this, a MANOVA was performed to determine if there were significant differences in ability (Gf and Gc) on the basis of age group membership. A 2 x 5 MANOVA was performed on the dependent variables: Verbal ability total score (V_{total}) and Abstract total score (A_{total}), of which were posited to represent Gc and Gf, respectively. All significant *F* values in the MANOVA were followed up by ANOVA analyses that treated A_{total} and V_{total} separately. These two variables are generally reported to correlate substantially (approximately .60), as such, a suitable follow-up procedure to scrutinize the pattern of relationships between the DVs and age was a Roy-Bargmann Stepdown Analysis (Mudholkar & Subbaiah, 1988). As determined a priori, the Gf variable (A_{total}) was entered into the model first. The reason for this is that measures of Fluid intelligence traditionally have a much stronger relationship with age across the adult life span than measures of Crystallised intelligence (Horn, 1989; Kaufman & Lichtenberger, 2006; Kaufman et al., 2008). In the Stepdown analysis an ANOVA was conducted first, with A_{total} as the DV and age as the IV. Following this, an ANCOVA was performed with V_{total} as the DV, age as the IV, and A_{total} as the covariate. Alongside the Stepdown procedure, standard univariate statistics were also calculated. A post-hoc analysis was performed using Tamhane correction. As a part of this, significant mean differences and data trends were identified.

The assumptions of the MANOVA and stepdown procedures were considered. Sample size per cell was excellent (ranging from 195 to 1375 respondents) and all observations were independent from one another. The univariate normality condition was not met on several occasions; however, the exceptional cell size theoretically safeguards the research from this violation (Tabachnick & Fidell, 2001). Univariate and multivariate outliers did not pose a significant problem. Although there were approximately 5 cases with considerable multivariate non-normality, the large sample size negated these potential negative impacts. There was significant heterogeneity of variances, however due to the large sample size this was not a significant issue (Tabachnick & Fidell, 2001).

Results

Internal consistency

A Cronbach's Alpha internal consistency analysis was performed on all selected items. Items that demonstrated low item-total correlations (a cut off of <0.200)—items that improved Cronbach's Alpha when deleted—were dropped from the respective item pools. Once dropped, the analysis was run again and further items that were now demonstrating insufficient corrected item-total correlation were removed. This procedure was repeated until all items on both scales were deemed to have item-total correlations greater than 0.200 and had content that was varied and sufficiently represented the construct. Table 18 in Appendix C illustrates the first analysis results for both scales. Table 3 below illustrates the summary of the procedure. Following the removal of items that did not meet the reliability criteria, the Abstract scale was left with 16 items, and the Verbal scale with 22 items. The final Cronbach's Alpha statistic was an acceptable 0.804 for the Abstract sub-test and an acceptable 0.796 for the Verbal sub-test.

Table 3. Reliability analysis: summary table

Step	Item Removal	Sub-scale	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
1	Full Items	Abstract	.791	.798	20
	Full Items	Verbal	.786	.783	30
2	3,4,6,14 Rem.	Abstract	.804	.806	16
	3,8,15,18,19,26,29 Rem	Verbal	.792	.794	23
3	*No change*	Abstract	*No change*	*No change*	*No change*
	Step 2 + 28 Rem.	Verbal	.796	.796	22

Note: Abstract sub-scale showed no further items to be removed after Step 2.

Descriptive analysis

Abstract sub-scale

In Table 4 the overall mean for A_{total} was 11.18 (*SD* = 3.326), with the maximum possible value being 16 (16 items). This shows a clearly negative skew. This negative skew value indicates that the mean Abstract score for all respondents was between 50% and 100% correct. The kurtosis value divided by the std. error of kurtosis is a statistically significant value of 4.51. In respect to individual Abstract item descriptive statistics, the only notable values are the high kurtosis values seen for AQ7 and AQ15. While these values are significantly higher than the other Abstract item kurtosis values, this alludes to the items of significantly lower difficulty (i.e., very easy items). It was decided that these two items would remain in the analysis as these were the only two Abstract items between 0.9 and 0.99 on the difficulty index.

Table 4. Abstract item descriptive statistics

	N	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Scale Total	5097	11.18	3.326	11.065	-.62	.034	-.311	.069
AQ1	4912	.59	0.492	0.242	-.371	.035	-1.863	.070
AQ2	5006	.65	0.477	0.227	-.635	.035	-1.598	.069
AQ4	4939	.68	0.468	0.219	-.756	.035	-1.430	.070
AQ5	5075	.72	0.449	0.202	-.977	.034	-1.046	.069
AQ7	5091	**.96**	0.185	0.034	-5.033	.034	**23.344**	.069
AQ8	5019	.57	0.495	0.245	-.287	.035	-1.918	.069
AQ9	5035	.58	0.494	0.244	-.324	.035	-1.896	.069
AQ10	5070	.75	0.432	0.186	-1.172	.034	-.628	.069
AQ11	4946	.52	0.499	0.249	-.098	.035	-1.991	.070
AQ12	4992	.75	0.431	0.186	-1.174	.035	-.623	.069
AQ13	5044	.89	0.315	0.099	-2.467	.034	4.090	.069
AQ15	5014	**.94**	0.23	0.053	-3.853	.035	**12.854**	.069
AQ16	4984	.80	0.399	0.159	-1.516	.035	.298	.069
AQ17	4778	.59	0.492	0.242	-.356	.035	-1.874	.071
AQ18	4823	.62	0.487	0.237	-.474	.035	-1.776	.071
AQ19	4582	.73	0.443	0.197	-1.043	.036	-.913	.072
AQ20	4612	.86	0.344	0.118	-2.112	.036	2.460	.072
Valid N (listwise)	3170							

Note. Excessive kurtosis values, emboldened, are representative of items with low difficulty. The associated high mean values reflect this.

Verbal sub-scale

In Table 5, the overall mean for V_{total} was 14.96 (*SD* = 3.887), with the maximum possible value being 22 (22 items). This also demonstrates a clearly negative skew. This negative skew value indicates that the mean Verbal score for all respondents was between 50% and 100% correct. The kurtosis value divided by the std. error of kurtosis is a statistically non-significant value of 1.44. In respect to individual Verbal item descriptive statistics, the only notable values are the high kurtosis figures seen for VQ1, VQ5, VQ6, VQ17, and VQ21. While these values are higher than the other Verbal item kurtosis values, this merely alludes to the items of lower difficulty (i.e., easy items).

Table 5. Verbal item descriptive statistics

	N	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Scale Total	5089	14.96	3.887	15.107	-0.62	.034	-.100	.069
VQ1	5083	**.88**	.326	.326	-2.328	.034	**3.420**	.069
VQ2	5024	.83	.371	.371	-1.804	.035	1.253	.069
VQ4	5051	.75	.431	.431	-1.178	.034	-.612	.069
VQ5	5075	**.90**	.304	.304	-2.619	.034	**4.861**	.069
VQ6	5072	**.90**	.299	.299	-2.687	.034	**5.220**	.069
VQ7	5058	.28	.449	.449	.976	.034	-1.048	.069
VQ9	5028	.84	.365	.365	-1.869	.035	1.492	.069
VQ10	4844	.65	.478	.478	-.615	.035	-1.622	.070
VQ11	4985	.38	.485	.485	.495	.035	-1.756	.069
VQ12	5074	.79	.406	.406	-1.438	.034	.069	.069
VQ13	5034	.35	.476	.476	.645	.035	-1.585	.069
VQ14	5061	.81	.396	.396	-1.546	.034	.391	.069
VQ16	5014	.65	.477	.477	-.635	.035	-1.598	.069
VQ17	5040	**.91**	.280	.280	-2.956	.034	**6.742**	.069
VQ20	4991	.78	.416	.416	-1.330	.035	-.231	.069
VQ21	4973	**.86**	.352	.352	-2.017	.035	**2.070**	.069
VQ22	4890	.65	.478	.478	-.620	.035	-1.616	.070
VQ23	4859	.47	.499	.499	.113	.035	-1.988	.070
VQ24	4669	.80	.403	.403	-1.472	.036	.167	.072
VQ25	4725	.68	.465	.465	-.794	.036	-1.370	.071
VQ26	4184	.33	.469	.469	.740	.038	-1.453	.076
VQ27	4501	.59	.491	.491	-.380	.036	-1.857	.073
VQ28	4292	.54	.499	.499	-.150	.037	-1.979	.075
VQ29	4152	**.92**	.271	.271	-3.110	.038	**7.679**	.076

VQ30	4028	.73	.443	.443	-1.053	.039	-.892	.077
Valid N (listwise)	3447							

Note. Excessive kurtosis values, emboldened, are representative of items with low difficulty. The associated high mean values reflect this.

The sub-test scores for each age group were also calculated. The means and standard deviations for both A_{total} and V_{total}, for each of the five age groups, have been calculated and presented in Table 6. In terms of Abstract total, there appears a somewhat linear negative relationship between age and mean score. The youngest age group (16-24 years) has a mean of 11.77 (*SD* = 3.09). This total score successively decreases up the age groups. The oldest age group (51-58 years) has a mean score of 8.12 (*SD* = 3.22).

Scale-total descriptives

In regards to the Verbal total score, there also appears to be a *slight* negative relationship between age and total score, however not to quite the same extent. The youngest age group has a mean of 15.25 (*SD* = 3.76). This total score decreases slightly up the age groups, to a mean of 13.22 (*SD* = 3.96) in the oldest age group. The standard deviations for all cells are relatively consistent with one another.

Table 6. Age group means and standard deviations for the Abstract and Verbal scales

	Sub-test			
	A_{total}		V_{total}	
Age Group (yrs)	Mean	Std. Deviation	Mean	Std. Deviation
16-24	11.77	3.09	15.25	3.76
25-33	11.68	3.06	15.34	3.79
34-42	10.88	3.35	14.63	3.85
43-50	9.60	3.61	14.18	4.25
51-58	8.12	3.22	13.22	3.96

Item difficulty analysis

Both scales' items were reviewed for their item difficulty. As part of this analysis, each item's difficulty level was calculated. In the present study, both the Abstract and Verbal scales differed on how discriminant they were. As can be seen in Figure 3, the item difficulty of the Abstract items is reasonably restricted. Ranging from 0.6 to 0.99—chance being 0.167—there is a veritable deficit in items ranging from 0.2 to 0.59. That is, it appears that there is a dearth of more difficult items (between difficulty levels of 0.20-0.59). As can be seen in Figure 4, the Verbal scale has a greater difficulty variance (between 0.30 and 0.99), with a slightly greater proportion of harder items.

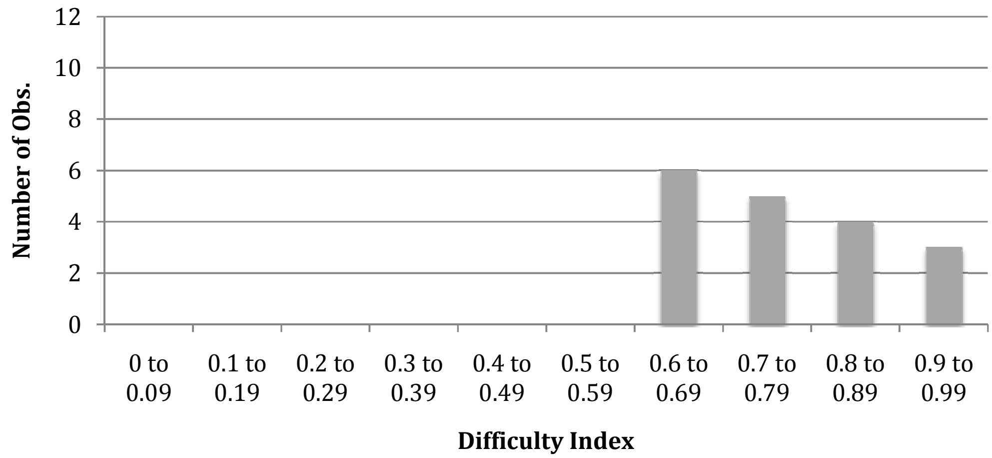

Figure 3. Abstract item difficulty index.

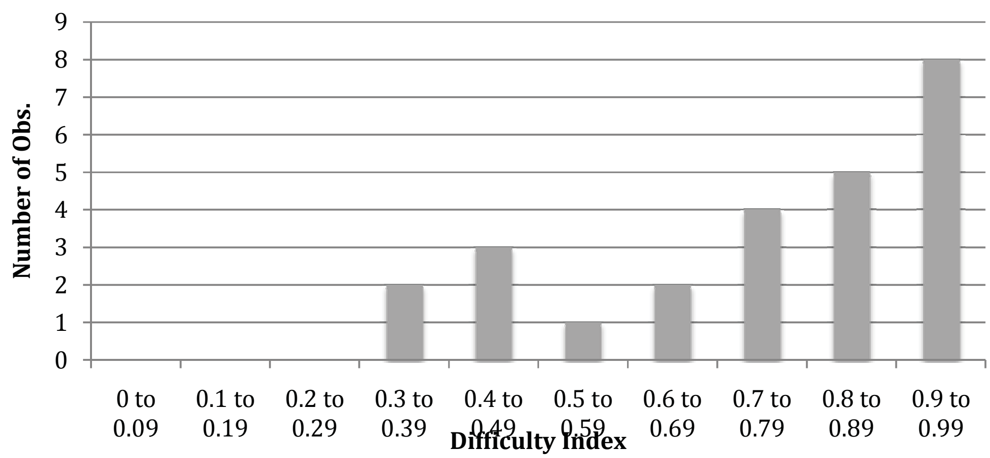

Figure 4. Verbal item difficulty index.

Confirmatory Factor Analysis: One-factor versus two-factor model

Using the AMOS statistics package (Ibm, 2009) confirmatory factor analysis was performed in order to identify whether a one-factor (general factor of intelligence, *g*) or a correlated two-factor model (Gf-Gc) would fit the data better. As suggested Table 7, all four of the model-fit statistics suggest better fit for the two-factor model. NFI was superior for the two-factor model, 0.876, versus a smaller 0.835 for the one-factor model; TLI was 0.894 versus 0.847, and CFI was 0.905 versus 0.862. The

RMSEA value for the two-factor model came out at 0.023 versus 0.028 for the one-factor model. The correlation for the two-factor model was a high 0.782. Overall, this suggests that out of the two proposed models, the hypothesized two-factor Crystallised-Fluid model is a better fit for the data than is a single factor "*g*" model. This finding further supports the notion that, given the current dataset, the Verbal and Abstract sub-tests are measuring two distinct constructs, and thus further supports the notion that they can be assessed as such in the subsequent analyses.

Table 7. One-factor versus two-factor model fit statistics

Model	Model-fit statistic					
	χ^2	df	NFI	TLI	CFI	RMSEA
1-factor	3348.036	665	0.835	0.847	0.862	0.028
2-factor	2507.465	664	0.876	0.894	0.905	0.023

Also of importance within these statistics is the *absolute*, not relative, fit of the two-factor model. With specific reference to CFI, the value is above the suggested acceptable level of 0.90. In considering this however, a deviation below this wouldn't have been detrimental to the fit of the model (Bentler, 1990).

<u>Group differences</u>

MANOVA

One of the prerequisites for MANOVA is the assumption of homoscedasticity. As can be seen in Table 8, the Box's M statistic is significant at the p < 0.001 level. Therefore, this assumption cannot be met. However, it is well known that the Box's M is highly sensitive to violations of normality and disparities in cell sample sizes. The current study contains a reasonable amount of univariate and multivariate non-normality and has hugely inconsistent cell sizes (ranging from n=195 to n=1929). The Box's M statistic will therefore be interpreted as such.

Table 8. Box's M statistic for equality of covariance

Box's M	41.492
F	3.452
df1	12
df2	5.28E+06
Sig.	0.000

Tested the null hypothesis that the observed covariance matrices of A_{total} and V_{total} were equal across groups.

Due to certain MANOVA requirements not being met, the more robust Pillai's Trace statistic was used (Tabachnick & Fidell, 2001). Table 9 shows the Pillai's Trace value was 0.076, which was a significant $p < 0.001$. The effect size of age was low (η^2=.038), accounting for only 3.8% of the variance in the combined DVs. This MANOVA indicates that there was at least one significant observed mean difference between two age groups on the composite ability variable (A_{total}-V_{total}). Figure 5 and Figure 6 illustrate the estimated marginal means for A_{total} and V_{total} across age. As can be seen both seem to have a slightly negative trend. Both the Gf and Gc representative scales appear to have a slightly negative linear trend. To further establish the degree to which each DV was contributing towards the significant difference, a Roy-Bargmann Stepdown analysis was performed.

Table 9. MANOVA: representing observed mean difference(s) in composite ability over age

DV		Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared
Age	Pillai's Trace	0.076	49.739	8	10124	0.000	0.038

a. The statistic is an upper bound on F that yields a lower bound on the significance level.

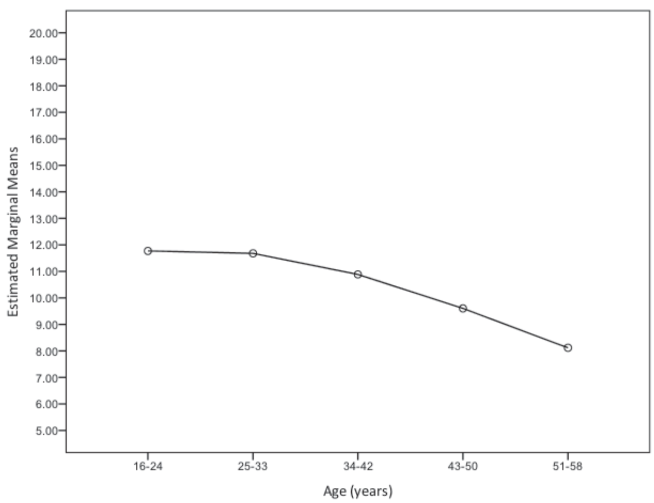

Figure 5. Estimated marginal A_{total} means for ages 16 through 58 years.

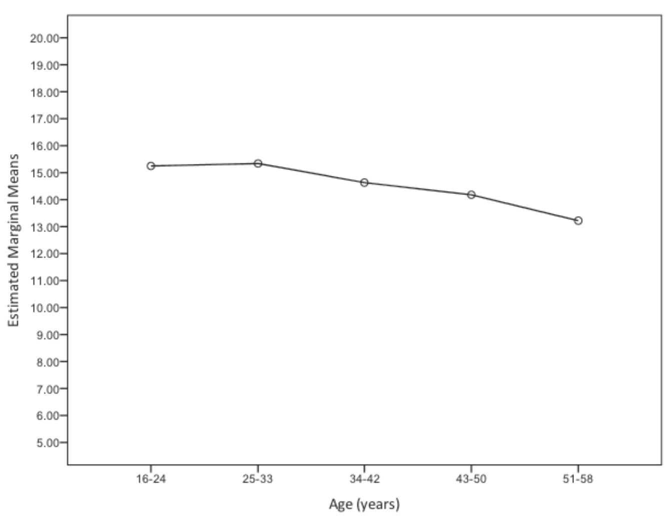

Figure 6. Estimated marginal V_{total} means for ages 16 through 58 years.

Post-hoc analysis

Post-hoc analyses were also conducted in order to determine how A_{total} and V_{total} varied among levels of age. The type of test conducted depended on whether variances of the groups were assumed to be equal. As seen in Table 10, the Levene's statistic suggested that variances of both A_{total} and V_{total} could not be assumed equal. Consequently, the Tamhane post-hoc test was used.

Table 10. Levene's test of equality of variances

	F	df1	df2	Sig.
A_{total}	10.808	4	5062	0.00
V_{total}	2.863	4	5062	0.02

Tested the null hypothesis that the error variance of A_{total} and V_{total} were equal across groups.

As suggested by Table 11, there is a negative relationship for both A_{total} and V_{total} as over increasing age. In regards to the Abstract scale, the only non-significant ($p > 0.01$) difference was between the age groups of 16-24 and 25-33 (0.092 units). The largest age group difference on the Abstract scale was between the youngest age group (16-24) and the oldest age (51-58). The difference was 3.653 units. Furthermore, the post-hoc results for the Abstract scale suggest a slightly increasing degree of ability decline between 16 to 58 years. The mean differences increase relatively linearly between the five age groups from 0.092 units (between 16-24 to 25-33 years) to 1.486 units (between 43-50 to 51-58 years).

In regards to the Verbal scale, there were three non-significant ($p < 0.01$) group differences. The first was between the 16-24 and 25-33 year groups (-0.091 units);

the second was between 34-42 and 43-50 years (0.452 units); and the third was between the 43-50 to 51-58 years (0.959 units). The largest age group difference on the Verbal scale was between the 25-33 year olds and the 51-58 year olds (2.118 units). The post-hoc results for the Verbal scale suggests that ability remains consistent between 16 to 24 years, drops slightly between 25 and 42 years, then stabilises between 43 and 58 years. This is not consistent with the curvilinear decline seen in the Abstract scale.

Table 11. Tamhane's post-hoc mean comparison

	(I) Age Group (years)	(J) Age Group (years)	Mean Difference (I-J)	Std. Error	Significance
A_{total}	16-24	25-33	0.092	0.109	0.99
		34-42	**0.886**	0.132	**0.00**
		43-50	**2.167**	0.184	**0.00**
		51-58	**3.653**	0.245	**0.00**
	25-33	34-42	**0.795**	0.123	**0.00**
		43-50	**2.075**	0.178	**0.00**
		51-58	**3.561**	0.241	**0.00**
	34-42	43-50	**1.281**	0.193	**0.00**
		51-58	**2.767**	0.252	**0.00**
	43-50	51-58	**1.486**	0.283	**0.00**
V_{total}	16-24	25-33	-0.091	0.133	1.00
		34-42	**0.617**	0.155	**0.00**
		43-50	**1.069**	0.218	**0.00**
		51-58	**2.028**	0.301	**0.00**
	25-33	34-42	**0.707**	0.145	**0.00**
		43-50	**1.159**	0.211	**0.00**
		51-58	**2.118**	0.296	**0.00**
	34-42	43-50	0.452	0.226	0.37
		51-58	**1.411**	0.307	**0.00**
	43-50	51-58	0.959	0.343	0.05

a. Based on observed means.

b. Emboldend values represent significance at the $p <$ 0.01 level

Roy-Bargmann Stepdown Analysis

Table 12 illustrates that age was a significant main effect for Gf and Gc, in both the stepdown analysis and the univariate ANOVAs. As was predicted, Gf demonstrated a larger effect size (η^2=.074) for the univariate ANOVA than did Gc (η^2=.018). The univariate analysis suggests that age explains 7.4% of the variance in V_{total} and 1.8% of the total variance in A_{total}. However, the stepdown analysis demonstrates that the unique impact of Age on V_{total} (i.e., independent of the association between Age and A_{total}) was a negligible 0.2% of explained variance. In Table 13, the adjusted V_{total} means for every age group appear similar.

Table 12. Summary of univariate and stepdown ANCOVA analyses on Gf and Gc

IV/Stepdown covariate	DV	Univariate *F (df)*	Effect Size	Stepdown *F (df)*	Effect size
Age	Gf	101.175** (4,5070)	0.074	101.175** (4,5070)	0.074
Age/Gf	Gc	23.507* (4,5062)	0.018	2.935 (5,5061)	0.002

Note: IV = independent variable; DV = dependent variable

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

Table 13. Estimated marginal means of V_{total} independent of A_{total}

Age	V_{total} Mean	Std. Error
16-24	14.85	0.085
25-33	15.00	0.072
34-42	14.85	0.095
43-50	15.29	0.144
51-58	15.37	0.229

Age-variance analysis

Finally, attention is brought to the discrepancies of group variances as initially demonstrated by the significant Levene's test. As seen in Table 14 there is a distinct difference in the pattern of variance between the Abstract scale and the Verbal scale. In terms of percentage of the mean, and with the exclusion of the youngest group, the standard deviations for the Abstract scale almost consistently increase with age. While variance in the 16-33 age range remains similar, from 33 years onwards it increases from 26.2% right through to 39.7%, an increase of 13.5%. Conversely, the variance for the Verbal scale (as percentage of the mean) ranges from 24.7% in the lowest two age ranges to only 30.0% for the 51-58 year age group, an increase of only 5.3%. While still a positive increase, it is a lower magnitude than the Abstract scale.

Table 14. A_{total} and V_{total} variance comparison across Age

		Age Group				
		16-24	25-33	34-42	43-50	51-58
A_{total}						
	Mean	11.77	11.68	10.88	9.60	8.12
	S.D.	3.09	3.06	3.35	3.61	3.22
	% of the mean	26.3	26.2	30.8	37.6	39.7
V_{total}						
	Mean	15.25	15.34	14.63	14.18	13.22
	S.D.	3.76	3.79	3.85	4.25	3.96
	% of the mean	24.7	24.7	26.3	30.0	30.0

Means and standard deviations as well as "percentage of the mean" ratios are presented.

Discussion

The primary aims of this research were; firstly, to evaluate the psychometric properties of the GRT2 measure, and secondly, to investigate the effects of age on intelligence, specifically in terms of Cattell and Horn's (1941, 1961, 1978) model of Crystallised and Fluid intelligence. Firstly, the GRT2 demonstrated adequate, but far from ideal, psychometric properties: Confirmatory Factor Analysis and reliability analysis results were satisfactory, but the item-difficulty analysis results showed room for improvement. Secondly, results suggest that age has a clear negative relationship with Fluid intelligence but not necessarily with Crystallised intelligence. Age was also accompanied by a distinct increase in the heterogeneity of variance in Fluid intelligence.

GRT2 psychometric evaluation

In the first component of this evaluation, the measure demonstrated acceptable internal consistency for both scales. Closer analysis suggested that the removal of certain items (4 from the Abstract scale and 9 from the Verbal scale), with low item-total correlations, could increase the alpha of the Abstract and Verbal scales used in the subsequent analysis. That these items could be removed to increase alpha, suggests that the items may have been selected during the test design phase without necessarily demonstrating sufficient internal consistency. It otherwise suggests that the item-level data used in the measure design process and the data seen in this study, demonstrated significantly different reliability properties.

Insufficient item-difficulty distributions were found for both the Abstract and Verbal scales. With regards to the Abstract scale, there appears to be an acute dearth of higher difficulty items. The Verbal scale also appears to have a deficit of higher-difficultly items, but not quite to the same extent. Ideally, there should be an item-difficulty distribution that has neither floor nor ceiling effects. As the data used in this analysis came from a norm group purported to be representative of the NZ population, it is concluded that the Abstract and Verbal scales, as used in current selection context, are comprised of too many easy items and not enough moderate and high difficulty items. This is a pertinent limitation of the measure in its use in employment selection procedures. The ceiling effect seen here, would have likely worked to decrease the amount of explainable variance for the effect of Age on both V_{total} and A_{total}.

The variance in Crystallised and Fluid intelligence explainable by age

The MANOVA and the subsequent Roy-Bargmann stepdown analysis suggests that Age has an overall significant negative semi-linear relationship with the measure of Fluid intelligence, but not necessarily with the measure of Crystallised intelligence. The peak for Fluid intelligence was indistinguishable between 16-24 years and 25-33 years. This is out of line with a profusion of previous research that suggests a distinct peak in the early 20s (Horn, 1985; Horn, 1989; Kaufman, 1990; Kaufman, 2001; Miller et al., 2009; Owens, 1966; Salthouse, 2004). It does however lend itself to an amalgam between the research that suggests this peak be in the 20s and with Kaufman et al.'s (2008) results, which allude to a peak in the 26 to 35 years age category. Overall, the Fluid ability aging pattern is evidently aligned with Horn and

Cattell's (1967) notion of Fluid intelligence as a "vulnerable" ability, and subsequent investigations thereof.

Crystallised intelligence, as measured by the Verbal scale, initially appeared to show a slight negative decline. However, upon controlling for Fluid intelligence, this relationship all but disappeared. Thus, Crystallised intelligence was found to have a non-significant relationship with age. This is in moderate disagreement with previous research, which seems to suggest that Crystallised intelligence increases steadily into the 60s, before declining slightly into the 70s (Cattell & Horn, 1978; Horn & Hofer, 1992; Horn & Cattell, 1967; Kaufman & Horn, 1996; Kaufman, 2001). That is, the Crystallised ability did not demonstrate the positive relationship with age that was predicted. Why these Crystallised ability results did not conform to the expected patterns cannot be explicated without further research. Though, it is suspected that three measure properties, in particular, may have impacted this finding. Firstly, designing items such that they load purely on to either Fluid or Crystallised ability is especially challenging. Secondly, given the timed nature of this assessment, the age-related decline in processing speed (Gs) (Mcgrew, 2005; Salthouse, 1985) may have attenuated the expected increase. Thirdly, the test may not have been robust enough, and may not have had enough item variability and item difficulty variability.

It is suspected that many of the Crystallised items (Verbal scale) may have been inadvertently loading onto Fluid ability. While this cannot be ascertained without further analysis, this suspicion arises from two observations: 1) on closer inspection, many of the Verbal scale items appear to involve a significant degree of relation eduction (i.e., identifying relations, and understanding inference on novel content) over knowledge requirements. As Crystallised ability items, they should be measuring one's level of informational awareness (Cattell, 1963). It could certainly be argued however, that the level of knowledge required to correctly answer many of the Crystallised items is particularly low. The ability to logically analyse the "level" at which say "time" and "month" (taken from a GRT2 Verbal item) are related to one another, and consequently use this information to infer the correct response, is arguably far more a Fluid-like ability. This type of item is somewhat prevalent in the GRT2 Verbal sub-test, thus it may be that the GRT2 Verbal scale is substantially contaminated by measures of Fluid intelligence. The second observation supporting this notion was the strong reduction in the explainable variance in Crystallised ability, upon the addition of Fluid ability as a covariate in the stepdown analysis. While the two abilities are indeed correlated (r=0.782 in the present study), the veritable abolition of the Age-Crystallised relationship upon controlling for Fluid intelligence suggests the Verbal item factor loadings may have been significantly less pure than initially suspected.

While it was endeavored that the Verbal measure be, as pure as possible, a measure of Crystallised intelligence, it was a well-considered fact that the GRT2 was designed and administered—for the current dataset—as a timed assessment. As opposed to a power assessment, the GRT2 requires that the testee focus on both speed and accuracy. Thus, not only is one's knowledge (Gc) invoked, but one's ability to process information efficiently is also measured, i.e., processing speed (Gs). This is a strong confound. That Gs has significant negative relationship with age (Botwinick, 1977; Mcgrew, 2005; Salthouse, 1985) could mean that the neutral Crystallised ability aging pattern seen in this study may have been further confounded. It is therefore speculated that a version of this assessment with looser time restrictions—everything else being equal—may have produced a Crystallised aging pattern more similar to what has been witnessed in previous research, i.e., maintained, but steadily increasing up until the 70s.

Not only are the changes in the relative levels of Crystallised and Fluid ability interesting when investigating the age-intelligence relationship, but also the age-related ability dispersion. An analysis of the scale variances revealed a distinctive difference in the pattern of variance between the measures representing Crystallised and Fluid intelligence. Unlike the Crystallised measure, Fluid intelligence's variance appeared to increase significantly and consistently over increasing age. That is, it appears that as people grow older, average person-to-person differences in ability to think logically and solve novel problems significantly increases. This finding is supported by Ardila (2007) who similarly found that the WAIS-III norm data demonstrated a significant increase in Fluid-like item score dispersions across age. This begs the question: what might be the reason for this increase in intellectual heterogeneity across age? If one were to hazard a guess, it could be that the increased variability is perhaps a product of educational opportunity and experience. The reasoning is that as people grow older, they have had more opportunity to engage in intellectual stimulation, and have therefore had more opportunities to attenuate the age-related degradation of neuro-physiological structure, an extremely tenuous hypothesis, but nonetheless seemingly plausible to this author.

Limitations

There are major limitations of this research that must be acknowledged when considering the aforementioned findings. Firstly, the analysis was an archival cross-sectional investigation on a dataset with limited availability of bio-data. As such, any potentially confounding cohort factors, such as educational attainment (which is suggested to have a strong relationship with Crystallised ability in particular (Kaufman & Lichtenberger, 2006)), have not been controlled for. This is likely to have had a reasonably significant impact on the present findings, the magnitude and direction of which, however, cannot be determined without further investigation.

Secondly, the archived dataset used in the current study lacked respondents in upper age-ranges (50+), so much so that the oldest age group had to be limited to 51-58. If this author had orchestrated the data collection, significantly more data would have been collected from respondents between the ages of 50 and 80 years. Unlike most of the WAIS standardisation studies, which have a roughly even distribution of candidates over the 20-90 year age range, the current investigation is limited to drawing conclusions only up to the 51-58 year age group. Had this data been available, significantly different aging patterns may have emerged.

Thirdly, the dataset has only limited generalizability in that the data was collected from respondents who were attempting employment for a “general level”. This sample may therefore differ from the New Zealand population in several ways. Those attempting employment at an organisation that requires ability testing, as opposed to those not a) seeking employment, or b) those seeking employment at an organisation that doesn’t require ability testing, may have a differing average level of ability as groups. Secondly, more intelligent individuals would, in general, be seeking employment in higher-level roles (not general level roles, for which the GRT2 was designed around). As such, it is likely that there is be noteworthy range restriction in the upper level of this distribution. This factor, combined with the insufficient difficulty range of the two scales suggests that the current findings have been further attenuated to some degree.

The fourth limitation is the lack of control that this research had over the collection of the archived dataset. While it is a contractual agreement between the test supplier and those administrating the assessments that it is done so in an appropriate and standardised manner, it cannot be known for sure to what degree this was practiced. Thus, an unknown quantity of error associated with administration consistency is coupled with this data collection.

Implications and suggestions for future research

While the GRT2 appears to work effectively as a predictive measure of likely job performance, its psychometric qualities, in terms of its ability to be deconstructed into a purely Crystallised-Fluid measure is lacking. The first piece of suggested future research would be a thorough item-level investigation into its factor loading and item design. Once the factor-loading pattern can be explicated (particularly the Verbal scale), the current research will be given significantly more detail in the context with which the results can be interpreted. Only then, can firmer conclusions be made regarding the Age-Verbal and Age-Abstract relations.

Furthermore, it is suggested that a similar study be conducted in the New Zealand context with a clinical based measure of cognitive Crystallised and Fluid intelligence, such as the WAIS-IV. The GRT2, as a commercial measure, is of little academic interest. As such, there is little to none peer-reviewed supporting research. Ideally, an investigation into Crystallised and Fluid intelligence in the New Zealand context would be based on a measure with which is easily accessible and of substantial standing in academia.

Reference list

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.

Adams, R., & Victor, M. (1993). *Principles of Neurology* (5th Ed. ed.): Mc-Graw Hill.

Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology, 22*(8), 1003-1011. doi: 10.1016/j.acn.2007.08.004

Baltes, P.B. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*(1), 12-21.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 201-213.

Bentler, P., & Bonett, D. (1980). Signifiance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 586-606.

Berowitz, B., & Green, R. (1963). Changes in intellect with age: Longitudinal study of Wechsler-Bellevue scores. *Journal of Genetic Psychology, 103*, 3-21.

Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins.

Birren, J.E., & Morrison, D.F. (1961). Analysis of the WAIS subtests in relation to age and education. *Journal of Gerontology, 16*, 363-369.

Boake, C. (2010). From the Binet-Simon to the Wechsler-Bellevue: Tracing the History of Intelligence Testing. *Journal of Clinical and Experimental Neuropsychology, 24*(3), 383-405.

Browne, M., & Cudeck, R. (1993). *Alternative ways of assessing model fit*. California: Sage.

Bugg, Julie M., Zook, Nancy A., DeLosh, Edward L., Davalos, Deana B., & Davis, Hasker P. (2006). Age differences in fluid intelligence: Contributions of general slowing and frontal decline. *Brain and Cognition, 62*(1), 9-16. doi: 10.1016/j.bandc.2006.02.006

Byrne, B. (1994). *Structural equation modelling with EQS*. California: Sage Publicaitons.

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

Catron, D., & Thompson, C. (1979). Test-retest gains in WAIS scores after four retest intervals. *Journal of Clinical Psychology, 35*, 352-357.

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

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

Cattell, Raymond B. (1987). *intelligence: Its Structure Growth and Action*: Elsevier Science Publishers B.V.

Cattell, Raymond B., & Horn, John L. (1978). A Check on the Theory of Fluid and Crystallized Intelligence with Description of New Subtest Designs. *Journal of Educational Measurement, 15*(3), 139-164.

Cole, Jason C., & Randall, Melissa K. (2003). Comparing the Cognitive Ability Models of Spearman, Horn and Cattell, and Carroll. *Journal of Psychoeducational Assessment, 21*(2), 160-179. doi: 10.1177/073428290302100204

Colum, R., Lluis-Font, J., & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutritional hypothesis. *Intelligence, 33*, 83-91.

Daniel, M. (1997). Intelligence testing: status and trends. *American Psychologist, 52*(10), 1038-1045.

Eisdorfer, C., & Wilkie, F. (1973). *Intellectual changes with advancing age*. New York: Springer.

Eysenck, H.J., & Fulker, D.W. (1979). *The Structure and Measurement of Intelligence*: Springer-Verlag, New-York.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Gregory, R.L. (1998). *The Oxford Companion to the Mind*: Oxford University Press, Oxford, UK.

Guilford, J.P. (1980). Fluid and Crystallized Intelligences: Two Fanciful Concepts. *Psychological Bulletin, 88*(2), 406-412.

Hernstein, R., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: The Free Press.

Horn, J. L. (1967). Intelligence - Why it grows and why it declines. [Article]. *Trans-Action, 5*(1), 23-32.

Horn, J.L. (1985). *Remodelling old models of intelligence*. New York: Wiley.

Horn, J.L, & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31*, 107-129.

Horn, J.L. (1989). *Cognitive diversity: A framework of learning*. New York: W.H. Freeman.

Horn, J.L., & Blankson, N. (2005). Foundations for Better Understanding of Cognitive Abilities.

Horn, J.L., & Cattell, R.B. (1982). Whimsy and Misunderstandings of Gf-Gc Theory: A comment on Guilford. *Psychological Bulletin, 91*(3), 623-633.

Horn, J.L., & Hofer, S.M. (1992). *Major abilities and development in the adult period.* Boston, MA: Cambridge University Press.

Horn, John L., & Cattell, Raymond B. (1966a). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253-270.

Horn, John. L., & Cattell, Raymond B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129.

IBM. (2009). AMOS Statistical Software.

Jensen, A. (1998). *The g factor*. Westport: Praeger.

Jensen, A.R., & Rushton, J.P. (2005). Thirty Years of Research on Race Differences in Cognitive Ability. *Psychology, Public Policy and Law, 11*(2), 246–248.

Jones, H., & Conrad, H. (1933). The growth and decline of intelligence: A study of a homogenous group between the ages of ten and sixty. *Genetic Psychology Monographs, 13*, 223-298.

Kaufman, A. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn and Bacon.

Kaufman, A. (1994b). *Practice effects* (Vol. 2). New York: Macmillan.

Kaufman, A., & Horn, J.L. (1996). Age changes on tests of Fluid and Crystallised intelligence for women and men on the Kaufman Adolescent and Adult Intelligence Test (KAIT) at ages 17-94 years. *Archives of Clinical Neuropsychology, 11*, 97-121.

Kaufman, A., & Lichtenberger, E. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

Kaufman, A., & Lichtenberger, E. (2006). *Assessing Adolescent and Adult Intelligence* (3rd ed.). Hoboken: John Wiley & Sons, Inc.

Kaufman, A., Reynolds, C., & McLean, J.E. (1989). Age and WAIS-R intelligence in a national sample of adults in the 20- to 74-year age range: A cross-sectional analysis with educational level controlled. *Intelligence, 13*, 235-253.

Kaufman, Alan S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence, 29*(2), 131-167.

Kaufman, Alan S., Johnson, Cheryl K., & Xin Liu. (2008). A CHC Theory-Based Analysis of Age Differences on Cognitive Abilities and Academic Skills at Ages 22 to 90 Years. *Journal of Psychoeducational Assessment, 26*(4), 350-381.

Kausler, D. H. (1982). *Experimental psychology and human aging*. New York: Wiley.

Kirkpatrick, E.A. (1903). *Fundamentals of child study: A discussion of instincts and other factors in human development with practical applications*. New York: Macmillian.

Kline, P. (1998). *The New Psychometrics: Science, Psychology, and Measurement*. London:UK: Routledge.

Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.

Lorge, I. (1936). The influence of test upon the nature of mental decline as a function of age. *Journal of Educational Psychology, 27*, 100-110.

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature, 297*, 222-223.

Lynn, R., & Vanhanen, T. (2006). *IQ and Global Inequality*. Augusta: Washington Summit Publishers.

Matarazzo, J.D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford.

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

McGrew, K. (2005). *The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future* (2nd ed.). New York: Guilford.

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

Miles, C., & Miles, W. (1932). The correlation of intelligence scores and chronological age from early to late maturity. *American Journal of Psychology, 44*, 44-78.

Miller, L. J., Myers, A., Prinzi, L., & Mittenberg, W. (2009). Changes in intellectual functioning associated with normal aging. *Archives of Clinical Neuropsychology, 24*(7), 681-688.

Ministry-of-Education. (2010). Ministry of Education: Learner summary. Retrieved 31st May 2011, from http://www.educationcounts.govt.nz/statistics/tertiary_education/provider_summary

Mudholkar, G., & Subbaiah, P. (1988). On a fisherian detour of the step-down procedure for MANOVA. *Communications in Statistics: Theory and Methods, 17*, 599-611.

Murphy, G. (1968). *Psychological views of personality and contributions to its study*. New York: Holt, Rinehart, and Winston.

Neisser, U., Boodoo, G. , Bouchard, T. J., Boykin, A. W. , Brody, N. , Ceci, S. J., . . . Loehlin, J. C. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(77).

Owens, W. (1966). Age and mental ability: A second adult follow-up. *Journal of Educational Psychology, 57*, 311-325.

Psytech. (Year unknown). *General & Graduate Reasoning Tests*. Bedfordshire: UK: Psytech International Limited.

Ryan, J., Sattler, J., & Lopez, S. (2000). Age Effects on Wechsler Adult Intelligence Scale-III Subtests. *Archives of Clinical Neuropsychology, 15*(4), 311-317.

Salthouse, T. (1985). *Speed of behavior and its implications for cognition* (2nd. ed.). New York: Van Nostrand Reinhold.

Salthouse, Timothy A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science, 13*(4), 140-144.

Sattler, J.M. (1982). Age Effects on Wechsler Adult Intelligence Scale-Revised Tests. *Journal of Counsulting and Clinical Psychology, 50*(5), 785-786.

Schaie, K.W. (1958). Rigidity-flexibility and intelligence: A cross-sectional study of the adult life-span from 20 to 70. *Psychological Monographs, 72*, Whole no. 462.

Schaie, K.W. (1983b). *The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood*. New York: Guilford.

Schaie, K.W. (1994). The Course of Adult Intellectual Development. *American Psychologist, 49*(4), 304-313.

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. New York: Macmillian.

Stankov, Lazar. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*, 35-51.

Sternberg, R., & Kaufman, J. (1998). Human abilities. *Annual Review of Psychology, 49*, 1134-1139.

Tabachnick, B., & Fidell, L. (2001). *Using Multivariate Statistics* (5th ed.). Boston: Allyn and Bacon.

Teasdale, T. (1989). Continuing secular increases in intelligence and a stable prevalance of high intelligence levels. *Intelligence, 13*, 255-262.

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

Valentin Kvist, Ann, & Gustafsson, Jan-Eric. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36*(5), 422-436.

Wang, Jing-Jen, & Kaufman, Alan S. (1993). Changes in Fluid and Crystallized Intelligence Across the 20- to 90-Year Age Range on the K-Bit. *Journal of Psychoeducational Assessment, 11*(1), 29-37.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Appendices

Appendix A: Abstract-Verbal as Fluid and Crystallised intelligence.

GRT2 Abstract scales as measures of Cattell's Gf

The GRT2 was designed as a measure of both Crystallised and Fluid intelligence (Psytech, Year unknown). There are multiple qualities of the assessment and results themselves, in which support this. In regards to the qualities of the assessment, Cattell and Horn (1967) list requirements of a measure to be thought of as a suitable measure Fluid and Crystallised intelligence. The three Fluid requirements are:

1. The assessments should be reflective of the eduction of relations and correlates and of an increasing hierarchical difficulty.
2. The assessments should base their fundaments of difficulty on higher-order non-verbal relations, such as part-whole relations, super-ordinate relations, developmental relations, and degrees of similarity.
3. The assessments should be free from any fundaments in which differential ability to understand the items stems from particular knowledge or cultural symbolism of some form.

In regards to the first requirement, the GRT2 Abstract sub-test requires the respondent to educe the most logical and parsimonious relationship between the series of visual-spatial figures. Furthermore, the difficulty index of the Abstract items range from 0.52 through to 0.96 (almost everyone responding correctly), demonstrating a moderate hierarchy of difficulty. In regards to the second requirement, the GRT2 Abstract component is comprised of 25 non-verbal visual-spatial items. Of which, the respondent is required to understand the complex visual relations and respond accordingly. In regards to the third requirement, all 25 non-verbal items of the GRT2 Abstract component are seemingly unbiased from any cultural symbolism or knowledge that may assist the respondent. All visual-spatial items were designed as to be completely abstract and non-meaningful (Psytech, Yeah unknown).

Based upon this, it is concluded that the GRT2 Abstract measure is sufficiently suitable as a measure of Cattell's Fluid intelligence.

GRT2 Verbal scales as measures of Cattell's Gc

According to Cattell and Horn (1978) and Carroll (1993), a measure of Crystallised intelligence is any assessment that makes informational demands, of which are considered skills or knowledge valued within the respondent's culture. The Verbal component of the GRT2 is comprised specifically of items that aim at measuring the extent of the respondent's vocabulary and of items that require respondents to deduce the relationship between words, thereby explicitly requiring the respondent to have in-depth knowledge and understanding of how, and at what level, the given words relate to one another.

Appendix B: Missing data analysis

Table 15. Missing data and item difficulty for Abstract items

Item	Valid N	Missing N	% Missing
AQ1	4912	185	3.77
AQ2	5006	91	1.82
AQ3	5087	10	0.20
AQ4	4939	158	3.20
AQ5	5075	22	0.43
AQ6	4765	332	6.97
AQ7	5091	6	0.12
AQ8	5019	78	1.55
AQ9	5035	62	1.23
AQ10	5070	27	0.53
AQ11	4946	151	3.05
AQ12	4992	105	2.10
AQ13	5044	53	1.05
AQ14	4859	238	4.90
AQ15	5014	83	1.66
AQ16	4984	113	2.27
AQ17	4778	319	6.68
AQ18	4823	274	5.68
AQ19	4582	515	11.24
AQ20	4612	485	10.52
AQ21	**4405**	**692**	**15.71**
AQ22	**4346**	**751**	**17.28**
AQ23	**3989**	**1108**	**27.78**
AQ24	**4022**	**1075**	**26.73**
AQ25	**3647**	**1450**	**39.76**

Note: emboldened items were removed from subsequent analysis.

Table 16. Missing data and item difficulty for Verbal items

Item	Valid N	Missing N	% Missing
VQ1	5083	14	0.27
VQ2	5024	73	1.43
VQ3	5035	62	1.22
VQ4	5051	46	0.90
VQ5	5075	22	0.43
VQ6	5072	25	0.49
VQ7	5058	39	0.77
VQ8	5060	37	0.73
VQ9	5028	69	1.35
VQ10	4844	253	4.96
VQ11	4985	112	2.20
VQ12	5074	23	0.45
VQ13	5034	63	1.24
VQ14	5061	36	0.71
VQ15	5045	52	1.02
VQ16	5014	83	1.63
VQ17	5040	57	1.12
VQ18	4996	101	1.98
VQ19	5007	90	1.77
VQ20	4991	106	2.08
VQ21	4973	124	2.43
VQ22	4890	207	4.06
VQ23	4859	238	4.67
VQ24	4669	428	8.40
VQ25	4725	372	7.30
VQ26	4184	913	17.91
VQ27	4501	596	11.69
VQ28	4292	805	15.79
VQ29	4152	945	18.54
VQ30	4028	1069	20.97
VQ31	**3851**	**1246**	**24.45**
VQ32	**3699**	**1398**	**27.43**

VQ33	**3563**	**1534**	**30.10**
VQ34	**3407**	**1690**	**33.16**
VQ35	**3107**	**1990**	**39.04**

Note: emboldened items were removed from subsequent analysis.

Appendix C: Reliability analysis

Table 17. Reliability analysis for Abstract scale: step 1

Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
AQ1	.290	.786
AQ2	.496	.774
AQ3	**.129**	**.795**
AQ4	**.190**	**.791**
AQ5	.300	.785
AQ6	**.134**	**.795**
AQ7	.242	.788
AQ8	.310	.785
AQ9	.406	.779
AQ10	.380	.781
AQ11	.450	.776
AQ12	.364	.782
AQ13	.331	.784
AQ14	**.179**	**.791**
AQ15	.226	.788
AQ16	.440	.778
AQ17	.461	.776
AQ18	.318	.784
AQ19	.459	.776
AQ20	.365	.782

Table 18. Reliability analysis for Verbal scale: step 1

Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VQ1	.207	.784
VQ2	.284	.781
VQ3	**.096**	**.787**
VQ4	.291	.781

VQ5	.313	.781
VQ6	.255	.783
VQ7	.410	.775
VQ8	**.126**	**.787**
VQ9	.240	.783
VQ10	.328	.779
VQ11	.419	.774
VQ12	.397	.777
VQ13	.260	.782
VQ14	.263	.782
VQ15	**.145**	**.786**
VQ16	.401	.775
VQ17	.297	.782
VQ18	**.182**	**.785**
VQ19	**.025**	**.791**
VQ20	.200	.785
VQ21	.309	.781
VQ22	.453	.773
VQ23	.451	.773
VQ24	.250	.782
VQ25	.288	.781
VQ26	**.178**	**.786**
VQ27	.213	.785
VQ28	**.127**	**.789**
VQ29	**.128**	**.786**