Cognitive abilities, personality and interests: Their interrelations and impact on occupation

Jason Timothy Major

Doctor of Philosophy in Psychology

The University of Edinburgh

2013

Declaration

I hereby declare that this work has been composed by me, and that it is my own work, except where it has been clearly indicated. Furthermore, the work has not been submitted for any other degree or professional qualification.

Jason Major

Acknowledgements

I would like to thank my supervisors, Dr. Wendy Johnson and Professor Ian Deary, for their great expertise and encouragement. During the PhD I benefited from an intellectually lively department, particularly in the area of Individual Differences. I would like to thank all those students and staff who contributed to it. I also would like to thank my parents and family for their support.

Abstract

Cognitive ability, personality and interests are three distinct topics of investigation for psychology. In the past two decades, however, there have been growing appeals for research and theories that address the overlap among these domains (Ackerman & Heggestad, 1997; Armstrong, Day, McVay, & Rounds, 2008). One example of such a theory is PPIK theory (intelligence-as-process, personality, interests, and intelligence-as-knowledge) by Ackerman (1996). Integrative theories have the potential of not only increasing our theoretical understanding of the development of these individual differences, but of and improving vocational guidance through better prediction of future occupation (Armstrong, Su, & Rounds, 2011; Johnson & Bouchard, 2009). The research of this thesis was centered on examining the links among cognitive ability, personality and interests. The data came from Project TALENT (PT), a nationally-representative sample of approximately 400,000 American high school students from 1960 (Flanagan et al., 1962). A secondary topic was whether an integrated view could improve the prediction of attained occupation. This was tested with occupational data from follow-up PT surveys, conducted 11 years after high school. The first study addressed the structure of the PT intelligence tests. Three popular models of intelligence were compared through factor analysis: the Extended Fluid-Crystallized (Gf-Gc), Cattell-Horn-Carroll (CHC) and Verbal-Perceptual-Image Rotation (VPR) models. The VPR model provided the best fit to the data. The second study was an investigation of linear and nonlinear intelligence-personality associations in Project TALENT. The ten PT personality scales were related to the Big Five personality factors through content examination, consistent with previous research (Reeve, Meyer, & Bonaccio, 2006). Through literature review of studies on intelligence and the Big Five, 17 hypotheses were made about linear associations and quadratic associations of personality traits with general intelligence (*g*). The majority of the hypotheses were supported in all four grade samples: 53% in male samples, and 58% in female samples. The most notable finding, contrary to previous research, was that quadratic associations explained substantive variance above and beyond linear effects for Sociability, Maturity, Vigor and Leadership in males, and Sociability, Maturity and Tidiness in females. The third study examined associations between cognitive ability and interests, and their capacity to predict occupational type. Specifically, Ackerman's PPIK theory suggests that there are two "trait complexes" that are combinations of cognitive abilities and interests (termed science/math and intellectual/cultural). Trait complexes were derived from PT data separately by latent class analysis and factor analysis. It was hypothesized that they should have validity equal to or greater than individual intelligence and interests scores in predicting attained occupation. Instead, trait complexes derived through latent class analysis predicted substantially less variance in occupation than individual scales. The factor-analytic trait complexes performed more like the scales, but one trait complex (which involved *g* centrally) was inconsistent with PPIK theory. Overall, the trait complexes of PPIK theory were not supported. The results of the three studies are discussed in the context of existing integrative theories, and suggestions for future research are provided.

Publications from this thesis

Major, J. T., Johnson, W., & Deary, I. J. (2012). Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and Extended Gf-Gc models. *Intelligence, 40*(6), 543-559.

Major, J. T., Johnson, W., & Deary, I. J. (in press). Linear and nonlinear relations between personality and general intelligence in Project TALENT. *Journal of Personality and Social Psychology*.

Major, J. T., Johnson, W., & Deary, I. J. (2013). *Trait complexes of cognitive abilities and interests and their relations to realized occupation*. Manuscript in preparation.

Table of contents (text)

Table of contents (Figures and Tables)

Chapter 1: Introduction

The two main domains of study in differential psychology are intelligence and personality. More recently, occupational interests have received increased attention, driven by the practical goal of improving vocational guidance. These three domains of individual differences are not entirely independent, but overlap (Ackerman & Heggestad, 1997; Barrick, Mount, & Gupta, 2003). This overlap has led researchers to call for an integrative theory of individual differences that takes intelligence, personality and interests into account (Ackerman & Heggestad, 1997; Armstrong et al., 2008).

The overall purpose of this thesis was to explore the relations among the intelligence, interests and personality domains. The secondary goal was to examine the potential for an integrated view to improve the prediction of attained occupation. In this introductory chapter, a review is provided of prominent theoretical models in the three domains. Research on their integration is then reviewed, and finally the ability of integrative theories to improve our understanding of occupational attainment.

The data used in this thesis came from Project TALENT (PT), a longitudinal and nationally-representative study of the aptitudes, interests, and backgrounds of American high school students, started in 1960. Chapter 1 provides background on how PT was conducted and the measures within it. Chapter 2 is a study that examined the structure of the PT intelligence tests, comparing three of the predominant intelligence models in the literature. Chapter 3 is a study that examined linear and non-linear associations between personality and intelligence. Chapter 4 is a study that focused on the associations between intelligence and occupational interests. It examined the character of potential "trait complexes" of intelligence and interests in the PT scales, and whether these trait complexes are better or worse predictors of future occupational type than individual scores for cognitive ability and interests. In chapter 5, a summary is given of what these studies have revealed about the intersection of intelligence, personality and interests, and the potential of integrative frameworks both to describe this overlap and be useful in the prediction of occupation.

1.1 Theories of intelligence, personality and interests

In order to study individual difference variables, psychologists need not only valid and reliable measures, but for multifaceted traits such as intelligence, personality and interests, theoretical models of their makeup. In this section, a brief but up-to-date picture is provided of the main theories for the three domains.

1.1.1 Intelligence

Research on the structure of intelligence has continued uninterrupted since the early twentieth century, when Spearman (1904) first proposed the concept of general intelligence (*g*), as the common factor underlying all cognitive ability tests. In the past several decades, research has converged on the hierarchical model as the best representation of cognitive abilities (Hunt, 2011; Reeve & Bonaccio, 2011). Here the term 'hierarchical' is used in the sense of a multiple-strata model, in which higher-order or more general cognitive ability factors are proposed to contribute directly to the lower-order or more specific ability factors. The lowest factor stratum consists of narrow abilities measured by individual tests. The second stratum consists of broad ability factors that emerge from higher-order factor analysis of narrow abilities. The third stratum emerges from factor analysis of broad abilities, but at this level only a single factor, known as *g*, is typically found.

Hierarchical models have both advantages and disadvantages in describing intelligence. One advantage is that the highly general concept of intelligence is divided into more manageable components called cognitive abilities (Reeve & Bonaccio, 2011). A cognitive ability can be defined as a latent trait that is observed from performance on particular cognitive tasks. Each ability is assessed by multiple tests, which vary in how purely they tap the ability (the remainder of test variance is made up specific test variance and cross-loadings on other factors that ideally are small in magnitude). A disadvantage of these models is that they are dependent to a certain extent on the properties of the tests in the battery, and on the testing sample (Hunt, 2011). In addition, subjectivity remains in interpreting to what the factors correspond at a more basic level, such as in a cognitive or biological sense (Hunt, 2010). An ultimate purpose of structural theories is to provide precise enough delineations of factors that their biological bases can be discovered, although this remains largely a future goal. Nonetheless, structural models have contributed greatly to advances in intelligence research, and are an essential part of current theories.

The three best-supported models in the intelligence literature are the extended fluid-crystallized (Gf-Gc) model (Horn & Blankson, 2005), the Cattell-Horn-Carroll model (CHC) model (McGrew, 2005, 2009), and the verbal-perceptual-image rotation (VPR) model (Johnson & Bouchard, 2005b). The differences among these models are covered in greater detail in chapter three, and are only summarized here. The models diverge primarily at the second stratum of the intelligence hierarchy. [1] The CHC model contains the greatest number of second-order factors: ten that have been firmly identified, and six more that have been characterized as "tentative" (McGrew, 2009). The extended Gf-Gc model contains eight second-stratum factors, which overlap strongly with those in the CHC model. This reflects the common origin of the two models, which can be traced back to Cattell's original fluid-crystallized model (Cattell, 1963), and Thurstone's primary mental abilities (Thurstone, 1938). The VPR model, by comparison, is more parsimonious and proposes only three second-stratum factors (the factors for which it was named).

The most notable difference between the second-stratum factors in the three models is that the CHC and Gf-Gc models contain factors which are delineated by how much they tap so-called fluid versus crystallized ability. Fluid intelligence refers to the ability to learn new information and solve novel problems, without regard to knowledge content or the content of material to which reasoning is to be applied, whereas crystallized intelligence refers to knowledge acquired from previous learning experiences (Cattell, 1963). These two factors are both present in the second-stratum of the CHC and Gf-Gc models. Moreover, the other factors can

[1] There has been some confusion in the literature surrounding the term stratum. Typically, this term has meant a level of a hierarchical model containing one or more factors. However, Johnson & Bouchard (2005b) characterized the VPR model as having four strata, counting the first level of individual tests as a stratum (p. 397). In the traditional sense, the model only has three strata. Reeve and Bonaccio (2011) also inaccurately presented the VPR model as having four strata.

be distinguished into those based more on “process” (Gf) compared to those based on “content” (Gc; Carroll, 1993). The VPR model, in contrast, posits that the second-stratum factors are distinguished only by their content. The factors in the VPR model are thought to be formed because the tests differ in the extent they are verbal (requiring the understanding of words and symbols), perceptual (requiring the understanding of visual-spatial stimuli), or image rotational (requiring the mentally rotation of visual-spatial stimuli (Johnson & Bouchard, 2005b)

Three factor-comparison studies have compared the three models presented, and in each case the VPR model displayed the best statistical fit (Johnson & Bouchard, 2005a, 2005b; Johnson, Te Nijenhuis, & Bouchard, 2007). However, these comparison studies relied on previous versions of the CHC model (the three-stratum model; Carroll, 1993) and the extended Gf-Gc model (the Gf-Gc model presented by Horn, 1998). Thus, a new comparison study was needed to distinguish among the three models. This is presented in the third chapter. In addition, the best-supported model was to be used in further examining associations with personality and interests.

1.1.2 Personality

The dominant model in personality psychology is the Five-Factor Model (FFM), which was first developed in factor analyses of personality trait terms by Tupes and Christal (1961; reprinted in 1992) and Norman (1963). However, the “Big Five” model only gained prominence in the mid-1980s after several different researchers found new empirical support for the model and argued for its theoretical merit (Goldberg, 1990; McCrae & Costa, 1985). Numerous labels have been put forward for each of the factors; however, the most commonly-used names were proposed by Costa and McCrae (1992): Extraversion, Neuroticism, Openness to Experience, Agreeableness and Conscientiousness.

The trait approach to personality has itself undergone many criticisms since its inception (Deary, 2009). Some have argued that the FFM is simply an empirical taxonomy, and thus that it lacks theoretical explanations for what the personality traits are, and how they emerge developmentally (Cervone, 2005; Cramer et al., 2012). Notwithstanding these more basic issues surrounding traits, the FFM has received considerable support as a taxonomic framework. The Five-Factor structure has been studied and partially replicated in over fifty cultures (McCrae & Terracciano, 2005, but see De Raad & Peabody, 2005). In addition, it has been found to capture the variance of personality factors on other major scales, such as the Eysenck Personality Questionnaire (Costa & McCrae, 1995), and Cattell's 16PF (Conn & Rieke, 1994, cited in McCrae, 2009).

There is the possibility that additional factors should be added to the Big Five. For example, Ashton and Lee (2005) proposed a sixth dimension termed Honesty-Humility, and up to eight broad factors have been supported by lexical studies (De Raad & Barelds, 2008). Nonetheless, these studies have also recovered the Big Five, supporting the position that they are "more-or-less sufficient to account for the co-variation of most personality traits" (McCrae, 2009, p. 148). Cramer et al. (2012) criticized the FFM for not being able to account for the variance in trait ratings without cross-loadings; however, they did not specify a hierarchical model (in the sense outlined above) that included facets. FFM proponents acknowledge that facet-level variance is a significant part of the FFM model. As Ashton and Lee (2012) observed: "Researchers have also known that a few broad factors can account for some large fraction of the covariation among personality variables, and not for all that covariation" (p. 433). Further refinement of the FFM at the facet and item level is still ongoing (McCrae, 2009).

The PT personality scales were not developed according to the FFM, but as described further in chapter 2, one study found a moderate level of correspondence between the PT scales and the Big Five (Reeve, Meyer & Bonaccio, 2006). The research on personality here was done in reference to the FFM, because it has proven a useful taxonomy for personality psychology. Moreover, the FFM has been used in much of the research aimed at discovering associations of personality with other individual difference variables (Ackerman & Heggestad, 1997; Barrick et al., 2003), and in the context of occupational prediction (Judge, Higgins, Thoresen, & Barrick, 1999). Thus the use of the FFM was helpful in forming hypotheses and relating the results back to the literature.

1.1.3 Occupational interests

Similar to the situation in personality psychology, one model is predominant in the occupational interests field: the RIASEC model (Holland, 1959, 1997). The model is composed of six broad interest factors: Realistic, Investigative, Artistic, Social, Enterprising, Conventional. The interest types are conceived as both manifestations of different work environments, and people's preferences for these environments (Holland, 1959, 1997). The six types are organized in a hexagon, in which the relations are expected to be highest between adjacent types, followed by alternative types (types separated by one in the hexagon), whereas opposite types are expected to have zero or negative associations. Consistent with these proposed associations, Prediger (1982) analyzed Holland interest scores for career groups and individuals and found support for two dimensions, named Data/Ideas and People/Things. These dimensions spanned opposite types: People/Things contrasted Social with Realistic interests, while Data/Things contrasted Conventional and Enterprising on one side with Investigative and Artistic on the other. Figure 1.1 Displays the Holland hexagon and Prediger's two dimensions. Hogan (1983) found two similar dimensions, which he called sociability and conformity, although the axes of these dimensions were rotated 30 degrees clockwise from Prediger's dimensions (Armstrong et al., 2011).

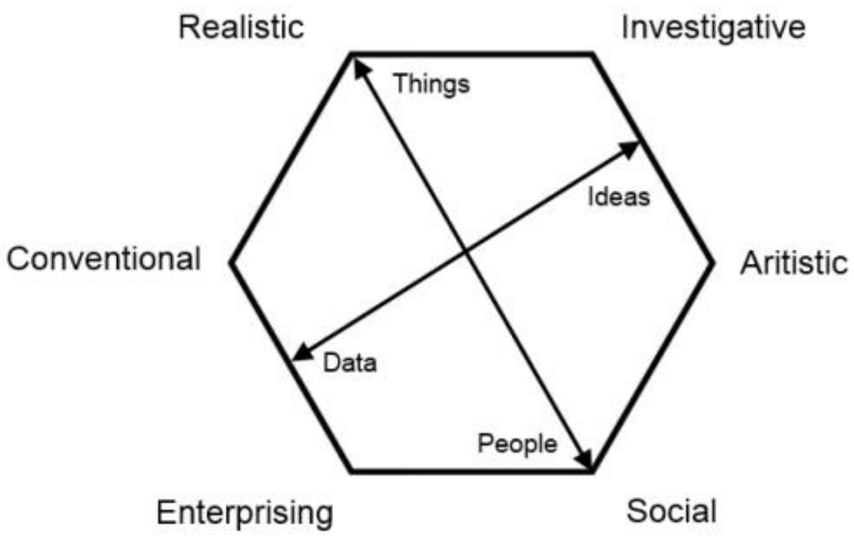

Figure 1.1 Holland's interests hexagon

In addition to these two dimensions that appear to underlie the RIASEC hexagon, a third dimension has been found termed prestige (Einarsdóttir & Rounds, 2000; Tracey & Rounds, 1996). This dimension was found when additional occupations were added to the scale underlying the RIASEC typology: the vocational preference inventory (VPI; Holland, 1985). These results suggested that the VPI has a restricted range of occupational prestige, an observation that was confirmed when a broader range of U.S. occupations was examined (Deng, Armstrong, & Rounds, 2007). In that same study, it was found that the prestige dimension was in fact not orthogonal to People/Things and Data/Ideas. The prestige dimension was associated with the Ideas pole of the Data/Ideas dimension. In addition, gender differences in occupational preferences were strongly associated with the People/Things dimension. Men tended to prefer jobs towards the Things pole, while women tended to prefer jobs towards the People pole, as had been observed in previous studies (Tracey & Rounds, 1992). Thus, the dimensions that underlie the RIASEC hexagon are related to prestige and gender differences, but the prestige dimension is only observed if a wider range of occupations than on the VPI is used (Deng et al., 2007).

To summarize, it appears that the RIASEC typology with its two dimensions is a reasonably adequate description of occupational interests for occupations, but this description could and probably should be expanded to give greater emphasis to a third dimension of prestige. The RIASEC model of interests is the primary one used in integrative research (Ackerman & Heggestad, 1997; Armstrong et al., 2008); thus it was an important point of reference for the research presented here.

1.3 Integrative theories

This section focuses on integrative theories in general, and does not cover all the literature on the overlap between interests, personality and cognitive abilities. For a review of cognitive ability and personality, see chapter 4. For a review of cognitive ability and interests, see chapter 5. The overlap of personality and interests was not reviewed because it was not addressed in the research here. There were two primary reasons for this. First, I suspected that cognitive abilities are the primary drivers of associations and so wanted to focus first on their associations. Second, it was necessary to keep the number of analyses in the study manageable (see chapter 5 for further discussion).

Psychologists have long hypothesized and observed that cognitive abilities, personality and interests are not entirely independent, but related; studies were conducted as early as Pearson in 1906 (see Ackerman & Heggestad, 1997, for a historical review). There are a number of possible reasons why a theory might be sought to explain these associations. Traditionally, however, the theories have most often been formulated in the context of explaining intellectual development (Ackerman, 1996). After Cattell first conceived of the concepts of fluid and crystallized intelligence (Cattell, 1943), the question emerged of how basic raw ability (Gf) developed into acquired knowledge (Gc). Cattell proposed the Investment Theory, which specified that Gc was the result of time invested, and of interest levels for specific areas of knowledge (Cattell, 1987). Vernon (1961) also theorized that industriousness and general academic interest both contributed positively to “educational ability”, but early studies were hindered by small sample sizes and a lack of broad measures for personality and interests.

As models in the three domains improved over time and measures become more standardized, it became increasingly possible to gather results from diverse studies. The first true meta-analysis of intelligence-personality associations was conducted by Ackerman and Heggestad (1997). This study also included a more qualitative review of the literature on interest-intelligence and interest-personality associations. These associations were the basis for a new theory of intellectual development called PPIK theory (intelligence-as-process, personality, interests, and intelligence-as-knowledge) by Ackerman (1996). The theory shared several similarities with investment theory, including maintenance of the distinction between raw or fluid ability (intelligence-as-process) and crystallized intelligence (intelligence-as-knowledge). Crucially, however, the theory specified that the overlaps among intelligence, personality and interests took the form of four “trait complexes”, which were defined as being similar to Snow’s concept of an aptitude complex in the learning domain (Snow, 1989). Snow proposed there were combinations of level of traits, such as cognitive abilities, personality traits and motivational traits that statistically interacted to produce better or worse outcomes in learning situations. Ackerman and Heggestad (1997) extended this concept to acquisition of academic knowledge more generally. Moreover, they claimed that this type of specialized knowledge is important to future occupation, a point addressed below.

A different approach to the integration of individual difference across the three big domains is to use occupational interests as an underlying framework (Anthoney & Armstrong, 2010; Armstrong et al., 2008; Armstrong et al., 2011). Armstrong and colleagues have proposed that the RIASEC model should be used because of its focus on work environments. They argued that educational and work environments are crucial to understanding the links among interests, personality and cognitive abilities, because these environments create demands for these traits, thus providing key contexts for them to become related (Armstrong et al., 2008). From the opposite perspective, it is thought that the demands of different occupational environments "pull" individuals towards them who have traits that would allow them to meet those demands. Thus, educational and work environments are thought to both have mutually reinforcing relations with traits, because they both select for the traits and potentially enhance them once individuals are in the environments. These ideas are not unique to Armstrong and colleagues, but have been proposed by a number of theorists on the development of occupational interests (Gottfredson, 2005; Hogan & Roberts, 2000; Scarr, 1996). One difference, however, is that Armstrong and colleagues made the specific claim that the RIASEC framework can be used to understand these relations.

Using a multiple-regression technique called property vector fitting, Armstrong and colleagues have attempted to fit personality traits and cognitive abilities onto the Holland hexagon (Anthoney & Armstrong, 2010; Armstrong et al., 2008). Armstrong et al. (2008) contained three studies; the first used data from several large studies that related the RIASEC types to personality traits (including the Big Five) and work styles from the Jackson Vocational Interest Survey (Jackson, 1977). A two dimensional RIASEC circumplex was specified, and property vector fitting was used to regress the personality and work style scores onto the two dimensions. Of the 51 personality traits and work styles, two-thirds (34) of them had more than 50% of their variance explained by the dimensions, indicating a good level of integration into the framework. The distribution of the traits provided support for both Prediger's (1982) and Hogan's (1983) underlying dimensions. In the second and third studies of Armstrong et al. (2008), cognitive ability was integrated into two and three dimensions of the RIASEC circumplex, which was also successful for a majority of the abilities. However, one limitation of these analyses as compared with Ackerman and Heggestad (1997) was that cognitive ability requirements were rated for different jobs, but were not derived from intelligence tests. Similarly, in Anthoney and Armstrong's (2010) study self-ratings of cognitive abilities were employed. Thus, there is a need to examine how actual cognitive ability scores fit into these models. Another limitation of Armstrong and colleagues' two studies was that the framework was dependent to a large degree on the RIASEC model, which likely does not give enough weight to job prestige (Deng et al., 2007).

A third possible integrative approach is to view interests and cognitive abilities as part of personality, broadly considered. DeYoung (2011) proposed that intelligence could be located underneath Openness to Experience in the FFM. In a previous study on the facets of Openness to Experience, DeYoung found that they were split into two domains: one labelled Openness which consisted of "aesthetically oriented traits", and the other called Intellect, which was formed from facets for intellectual engagement or self-perceived intelligence (DeYoung, Quilty, & Peterson, 2007). General cognitive ability was most strongly correlated with the Intellect aspect of Openness to Experience, as represented by the Ideas facet on the NEO PI-R (DeYoung, Peterson, & Higgins, 2005). However, this is a simplified picture because *g* has shown many smaller relations with at least one facet for each of the Big Five (DeYoung, 2011). In addition, narrower cognitive abilities beyond *g* are likely to have differential relations with personality traits. For example, Ackerman and Heggestad (1997) found that Conscientiousness is associated with Conventional interests in the RIASEC, which are in turn related to Perceptual Speed; thus it would be predicted that Consciousness is also associated with Perceptual Speed, although this association has not yet been observed directly.

From DeYoung's (2011) viewpoint, occupational interests fit within personality at the level of characteristic adaptations. Characteristic adaptation is a concept taken from the personality theory of McCrae and Costa (2012); it is defined as an acquired attribute, such as skill or attitude, that arises from the transaction of the person with the environment. Characteristic adaptations are contrasted with the basic tendencies that underlie the Big Five, which are thought to be more biologically-based and resistant to environmental influence. In apparent opposition to this view, behaviour genetic research has found that vocational interests display similar heritability coefficients to personality traits (Lykken, Bouchard, McGue, & Tellegen, 1993). However, Lykken et al. (1993) suggested that much of the heritability of interests could be explained as an indirect effect of genetic influence on other attributes, such as physique, personality, and cognitive ability. The heritability of interests could also be the result of gene-environment interaction and correlation of more basic traits; for example, if personality affects the initial selection of learning environments, and the success of individuals in those environments (Lykken et al., 1993). This hypothesis is echoed in a number of investment theories (Bouchard, 1997; Gottfredson, 2005; Hogan & Roberts, 2000; Scarr, 1996). Nonetheless, a major disadvantage of theories explaining occupational interests from this perspective is that they do not contain the detailed predictions of trait overlap that are provided in Ackerman's PPIK theory and Armstrong's framework.

The three integrative theories of personality, interests and cognitive abilities can potentially be distinguished by examining how well their models of the overlap match empirical data. PPIK theory proposes that this overlap is characterized by four trait complexes that involve groupings of high levels of particular personality traits, cognitive abilities and interests. The framework of Armstrong and colleagues instead suggests that personality and cognitive abilities should be mapped as continuous variables onto the RIASEC model of interests (Armstrong et al., 2011). The interests are the primary focus because they refer to preferences for education and work environments, which are theorized to be the contexts in which cognitive abilities and personality become related to each other and to interests. Finally, DeYoung (2011) has theorized that cognitive abilities and interests can be integrated into the FFM model of personality, where cognitive ability is found primarily under

Openness to Experience, and interests are characteristic adaptations resulting from the transaction of personality traits and cognitive abilities with the environment.

The last study of this thesis was focused first on examining the content of ability-interest trait complexes of PPIK theory. There were two main reasons for selecting PPIK theory. First, the proposed trait complexes were more parsimonious and specific than the many possible overlaps between cognitive abilities and interests in Armstrong's and DeYoung's approaches. This made them easier to identify and potentially falsify. Second, Ackerman and colleagues put forth the hypothesis that the trait complexes would demonstrate better predictive validity for occupation than individual scores for the three trait domains. For example, Ackerman and Beier asked: "is there a synergy among elements within the trait complexes, so that concentrating on trait complexes is more informative in the career choice context than individual trait measures?" (2003a, p. 209). This question provided another prediction of PPIK theory to test.

1.3 Prediction of occupational type

The three integrative approaches presented thus far have been assessed on their merits as theoretical frameworks, but a key issue is how they could potentially improve our ability to understand and predict occupational attainment. While the other theories have not involved as strong a claim for predictive power as PPIK theory, predicting occupation is a stated goal for most research in this area (Armstrong et al., 2008). As cognitive ability and personality are both related to and predict occupation (De Fruyt & Mervielde, 1999; Schmidt & Hunter, 2004), it is logical to hypothesize that integrative theories could provide superior prediction to considering each of the domains separately.

PPIK theory has a notable advantage over the frameworks of Armstrong et al. (2008) and DeYoung (2011), in that it has existed for a longer time, and thus more research has been done to link the theory to real-world outcomes. Ackerman and colleagues have found that their trait complexes relate to academic knowledge (Ackerman & Rolfus, 1999), university course selection (Ackerman, 2000) and university course performance (Kanfer, Wolf, Kantrowitz, & Ackerman, 2010). The results were taken as support for PPIK theory because knowledge is hypothesized to act a mediator between trait complexes and occupational attainment (Ackerman, 1996). Nevertheless, these studies only provide indirect evidence for theory because the predictive validities of either trait complexes or knowledge have not been examined for attained occupation.

In contrast to the indirect evidence for PPIK theory, there is not yet any evidence that the approach of Armstrong and colleagues could improve the prediction of occupation. This would require research relating the framework to occupational outcomes, possibly comparing its predictive validity to other theories. Previous research has demonstrated that personality and cognitive abilities can be mostly effectively integrated into two or three RIASEC dimensions (Anthoney & Armstrong, 2010; Armstrong et al., 2008). As the RIASEC dimensions are linked closely with preferences for different educational and occupational environments (Holland, 1997), the model with these dimensions could be useful in predicting future occupation, but this remains hypothetical.

DeYoung (2011) has provided a theoretical argument for how cognitive abilities and interests can be fit into the FFM. However, this account remains very general and does not specify, for example, which personality traits are involved in the formation of which occupational interests, or how narrow cognitive abilities fit into the FFM. Without these details it is not yet possible to use this theory to predict occupation.

Of the three integrative theories in the literature, PPIK is the most developed. It has made the most specific predictions for the overlap between cognitive abilities, personality and interests, and some indirect evidence has been found that trait complexes relate to occupation. For these reasons, I chose to test this theory in Project TALENT.

Chapter 2: Project TALENT's design and measures

Project TALENT (PT) was a study approved by the U.S. Department of Education in 1959 (Flanagan, et al., 1962). It was first conceived by John C. Flanagan, a professor of psychology at the University of Pittsburgh, who became the principal investigator. It was designed to be a longitudinal and nationally-representative study of the human talent of high school students, examining how this talent could be better identified and promoted. For example, the U.S. Commissioner of Education stated that the project was "an attempt to determine why so much of the nation's human potential is lost and what schools, counselors and parents can do to reduce the loss" (p. 1, Flanagan, 1962). To this end, a large amount of information was to be collected about the students (e.g. their aptitudes, interests and social backgrounds), as well as schools (e.g. their resources and teaching methods). The following description of the study relies heavily on the first PT report (Flanagan, et al., 1962). Details of the testing materials are also provided in the Project TALENT handbook (Wise, McLaughlin, & Steel, 1979). The computerized PT data was compiled by the American Institutes for Research, a nonprofit social science research institute founded by Dr. Flanagan. The data is available through the National Archive of Computerized Data on Aging (NACDA), from which they were obtained for the current research.[2]

The advisory panel of Project TALENT and its staff designed the study and its measures in 1958 and 1959. The schools were selected using a stratified random sample of public and private high schools across the United States. In all, 1353 schools were eventually sampled (93% of those asked), and approximately 440,000 students, who represented approximately 5% of the total American high school population. To enable the administration of the tests in each area, 90 regional coordinators were employed. The regional coordinators were primarily psychologists who were asked to work with local school administrations and teachers. Teachers and guidance counselors were trained to administer the Project TALENT tests, which they gave over two days. The initial testing occurred in

[2] The website of the NACDA is http://www.icpsr.umich.edu/icpsrweb/NACDA

March of 1960. Follow-up mail surveys were conducted after the students completed high school, after the intervals of 1, 5 and 11 years. The follow-up surveys asked about the participants' personal, educational and career experiences. Most relevant to the study presented in chapter 5, participants were asked about their current occupations at those times. In this chapter the data used in the thesis are first described: the measures of intelligence, personality, occupation interests, and follow-up occupational status.

2.1 Intelligence tests

The PT aptitude and achievement tests were newly-designed for the study. Their stated purpose was to "survey a variety of human aptitudes and to obtain scores which might predict an individual's ability to develop those aptitudes for vocational and educational success" (p. 57, Flanagan et al., 1962). One of the main reasons that new tests were created is that it was felt that pre-existing intelligence tests did not survey a wide enough variety of aptitudes, partly because the individual subtests were too long, and the ones in PT should be shorter. In addition, this would make it certain that none of the students had been previously exposed to the new tests.

A first experimental battery of all the tests was given to a sample of approximately 6000 high school students, in schools in the Northeast, South and Midwestern U.S. (Flanagan, et al., 1962, p. 60). Item-level analysis was used to exclude items that were unreliable, too hard, or too easy. Following this process, the final 60-test version was developed. The battery was composed of two main sections: the information tests, and the specific aptitude and achievement tests. For detailed test descriptions see chapter 3; here their general purpose and design is outlined.

The information tests were multiple-choice knowledge questions on a very broad range of topics, including both general knowledge and academic subjects. There were 36 subtests that ranged from 2 to 24 items in length. There were several purposes to the information tests. Firstly, it was held that the breadth of a person's knowledge was a measure of general intelligence; similar information tests were used in this way in the Army Alpha and Otis Mental Ability batteries. Second, the more specific tests had the potential to capture achievement in particular areas, as well as interest and motivation towards those topics, such as physical science, fine art or sports. Third, there was a vocabulary scale, which was regarded as a measure of verbal intelligence. In practice, the usefulness of many of the smaller information subscales was limited because of their narrow topics and poor reliability (Cureton, 1968; Flanagan et al., 1964). Cureton (1968) recommended that tests with less than nine items be excluded for intelligence research, which eliminated 15 tests. In addition to this, there were a number of tests that were highly likely to be sex-biased. For example, tests of Sports and Farming information favoured boys, whereas the Home Economics tests required knowledge to which girls were more likely to be exposed. Avoiding unreliability and sex-bias meant that only a maximum of 16 out of 36 tests (44%) were used in the studies of this thesis.

There were 24 aptitude and achievement tests in the PT battery. The types and number of aptitude tests were as follows: verbal (3), spatial visualization (2), reasoning (3), memory (2) and processing speed (4). The achievement tests included five English tests and three Math tests. The English tests assessed basic writing and reading skills learned in school, and the Math tests assessed arithmetic, introductory Math (studied in 9^{th} grade) and advanced Math (studied in grade 10 or later). The advanced Math test was left out of all studies here because it was deemed to be unfair for younger students. Generally, all of the aptitude and achievements tests were used in assessing cognitive ability in the present research, because most did not require prior knowledge, and even those that did (e.g. the English tests), only demanded basic knowledge to which all students should have been exposed.

2.2 Personality tests

There were 150 personality items in PT, which were in a section entitled "Student Activities Inventory". Students were asked to respond to statements about behaviors or characteristics in terms of how well they described "the things I do and the way I do them". Reponses were on a five-point Likert scale. The scores on the items were summed to form 13 scales; however, three of the scales were experimental and were not electronically recorded. Thus, there were ten personality scales made up of 108 items, with 7 to 24 items per scale. Item-level data were not available from the PT dataset, only scale scores.

The scales assessed general personality, but were aimed at personality traits that were important in educational and occupational contexts. For example, Flanagan et al. (1962) stated that “the TALENT [personality] battery was based on the hope that it would eventually add to our knowledge of how personality differences help to account for the differences in accomplishments of equally talented normal people” (p. 130). Thus, the scales have strong representation of traits that would be classified under Conscientiousness in the Big Five. However, Reeve, Meyer and Bonaccio (2006) re-administered the PT items to 219 university students, and found that the scales spanned all of the Big Five (as assessed by the NEO-PI-R). In a joint factor-analysis with the NEO, each of the Big Five received at least one substantial loading (mean r = .70, range = .51 to .81) from the PT scales.

Another difference of the PT personality scales from conventional personality assessment is that the students were aware that the purpose of the study was to examine talent, and it was conducted in a school context. Thus, although the instruction to students was to reflect on their general behavior, the context may have influenced them to respond in a manner more consistent with how they perceived themselves within school, or how they wished to portray themselves in a school setting. This possible confounding factor is explored in greater detail in chapter 4.

2.3 Occupational interest tests

The Interest inventory of PT was composed of 205 items, of which 122 were occupation titles and 83 were occupation-related activities. Students were asked to indicate how much they would like to do the occupation or activity. The PT study designers evaluated pre-existing interest scales, such as Strong’s Vocational Interest Blank and Kuder’s Preference Record, but decided to construct new scales.

The interest items were compiled into seventeen scales by PT investigators, based on a priori classification of different occupational areas. Using fifteen independent raters, Reeve and Hakel (2000) found that all of the PT interest scales except one (Labour) could be assigned to the RIASEC categories with acceptable accuracy (inter-rater agreement of 66% or higher). Unlike the personality scales, however, item-level data were available for the interests. As the original PT interest scales were not created on an empirical basis, in chapter 5 new scales were derived by factor analysis.

2.4 Occupation at follow-up

The PT follow-up data were collected primarily by the use of mail questionnaires. The questionnaires were designed to give a broad overview of the participants' lives after high school, thus in addition to questions on educational and occupational experience, they were asked about their marital status, quality of life, health, and other social variables (Wise et al., 1979).

Of the greatest relevance to the current research were the 11-year follow-ups, which occurred in 1971 to 1974, when the participants were approximately 28 years of age. Current occupation was asked in a written response, which was originally transformed into over 1000 occupation codes. These specific codes were later reduced to 254 job codes representing specific jobs or job areas such as Airplane Navigator, Veterinarian or Metal Trades (Wise et al., 1979). The job titles were also organized into twelve categories according to broad occupational themes. Greater detail on the occupation categories and the frequencies of participants in each are provided in chapter 5.

Although efforts were made to contact all PT participants, the participation rates for each subsequent follow-up decreased. Much of this attrition was due to lack of the most recent addresses for participants; addresses were lost for approximately 5% of participants for each year, in each grade, compared to baseline. Response rates also decreased with the time between the baseline testing and follow-ups, which were longer for the participants who were in lower grades (younger) at baseline. For the 11-year follow-up, 28.8% of the grade-12 participants returned the questionnaires, but only 19.6% of the grade-9 sample. To deal with attrition and the lack of representativeness of the follow-up samples, PT investigators conducted special interviews with approximately 2500 non-respondents to the mail questionnaires. The missing participants were found by a variety of methods, such as searching telephone directories, asking the Department of Motor Vehicles, and contacting the high school for new addresses. Once participants were located they were given telephone or in-person interviews (Wise et al., 1979). Sample weights were created in accordance with the sampling ratio of the special sample to original the 1960 sample (Wise et al., 1979). These sampling weights could then be used to adjust the follow-up sample to be representative of the baseline sample. The sampling weights were used in chapter 5 when follow-up occupation was being investigated.

Chapter 3: Comparing the VPR, CHC and Extended Gf-Gc models

3.1 Introduction

Disagreement about the structure of intelligence has a long history in psychology. Recently, however, some researchers have proposed that a consensus theory has emerged in the form of the Cattell-Horn-Carroll (CHC) theory of cognitive ability (Benson, Hulac, & Kranzler, 2010; Flanagan, Ortiz, & Alfonso, 2007; McGrew, 2005, 2009). McGrew (2005), for example, asserted that: “[Carroll’s synthesis] has finally provided both intelligence scholars and practitioners with the first empirically-based consensus Rosetta stone from which to organize research and practice” (p. 171). This view was contradicted, however, by three recent studies in which an updated version of Vernon’s verbal-perceptual model (1961, 1965) was found to provide better fit to large intelligence test batteries than the two precursors of the CHC model: Horn and Cattell’s fluid-crystallized (Gf-Gc) model (Cattell, 1963; Horn & Noll, 1997) and Carroll’s (1993) three-stratum model (Johnson & Bouchard, 2005a, 2005b; Johnson, Te Nijenhuis, et al., 2007). Nevertheless, these previous comparison studies relied on an interpretation of Gf-Gc theory which only included fluid and crystallized ability as second-order factors, whereas the Extended Gf-Gc theory contains six more such factors (Horn & Blankson, 2005). In addition, CHC theory has largely supplanted the three-stratum theory, and contains a number of differences from it (see below, and McGrew, 2009, for details). The current study was thus aimed at providing an updated test of whether the verbal-perceptual-image rotation (VPR) model (Johnson and Bouchard, 2005a), the CHC model, or the Extended Gf-Gc model provides a better description of the structure of intelligence.

Deciding among these models is an important issue for intelligence researchers because each implies a different underlying theory about the nature of intelligence and its manifestation. The three models have also each received substantial empirical support (Carroll, 1993; Horn & Noll, 1997; Hunt, 2011; Johnson & Bouchard, 2005b). The fluid-crystallized model has arguably been the most influential theory of intelligence to date in terms of the frequency of its application in research and test development (Carroll, 1993; Horn, 1998; Mackintosh, 2012). As Kaufman (2012) recently observed "the core concepts of *Gc* and *Gf* are still universal to nearly all IQ tests" (p. 119). It has also been claimed that the CHC model has the most cumulative factor-analytic evidence supporting it, thanks in large part to Carroll's (1993) major synthesis (McGrew, 2009). Although the VPR model has not been as prominent in the literature as the other two, we argue below that it has a number of advantages over the Gf-Gc and CHC models.

The main features of the three models are outlined in Table 3.1. These features are highlighted because they best reflect the theoretical differences among the models. Although these differences are based upon the most recent versions of the models, they have their roots in longstanding disagreements about the structure of intelligence. The CHC and Gf-Gc models are products of the American school of intelligence research, while the VPR model has its origins in the British school (Vernon, 1961, Carroll, 1993). In the early twentieth century, the divergent views of these two schools on the structure of ability were represented by Spearman and Thurstone. Spearman and his fellow British psychologists such as Burt (but not Thomson) emphasized the importance of the general factor of intelligence (*g*) over group factors in the structure of cognitive ability, whereas American psychologists, led by Thurstone, supported a model of orthogonal group factors, named primary mental abilities, with no general factor (Thurstone, 1938). Spearman and his colleagues argued that Thurstone's seven to nine primary factors were correlated and thus could also yield a model with a general factor and smaller group factors (Eysenck, 1939; Speaman, 1939). Whereas Thurstone rather quickly acknowledged the presence of higher-order factors in his datasets (Thurstone, 1947, cited in Carroll, 1993), and helped to develop the techniques for higher-order factor analysis, his reluctance to accept Spearman's *g*, and his conception of independent primary mental abilities had a lasting influence upon American intelligence researchers (Carroll, 1993). Notably, Cattell and Horn followed Thurstone in not accepting a *g* factor in their Gf-Gc model (Horn & Noll, 1997). In contrast, the g factor was prominent in Vernon's verbal-perceptual model (1961), and remains so in the VPR model. Nevertheless, it should be noted that VPR theory is agnostic about whether *g* represents a reflective or formative variable. We took the latent factor model approach to the VPR model in the current study, yet the model could be reformulated to conform to other approaches to the positive manifold (D. J. Bartholomew, Deary, & Lawn, 2009; Van Der Maas et al., 2006).

Table 3.1
Primary features of the CHC, Extended Gf-Gc and VPR models.

Feature	CHC model	Extended Gf-Gc model	VPR model
g factor postulated?	yes	no	yes
Number of second-order factors	10 (plus 6 more "tentatively identified")	8	3
Second-order factors are distinguished as content factors versus raw ability factors, or by content only.	Content (Gc, Gq, etc.) versus raw abilities (Gf, Gsm, etc.) [a]	Content (Gc) versus raw abilities (Gf, Gs, etc.) [b]	Content only
Number and nature of first-order factors	Pre-specified	Pre-specified	Left to battery content

[a] Gq is quantitative knowledge, Gsm is short-term memory (see McGrew, 2009).
[b] Gs processing speed (McGrew, 2009).

Although the CHC model does contain a *g* factor, its second-order factors are highly similar to those in the Gf-Gc model. This is because the CHC model was formed by merging Carroll's (1993) three-stratum model with the Gf-Gc model (McGrew, 1997, 2005). Carroll himself was also strongly influenced by Gf-Gc theory, writing that prior to his theory it was "the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities" (p. 62, 1993). The original Gf-Gc model had only two second-order factors of fluid and crystallized ability; however, Cattell and his student Horn eventually added six other second-order factors. These latter factors resemble Thurstone's primary abilities: for instance, quantitative knowledge (Gq), which is similar to the primary ability numerical facility, visual processing (Gv), which is similar to the primary ability spatial relations, and processing speed (Gs), which is similar to the primary ability perceptual speed (Horn & Blankson, 2005; Mackintosh, 2012; McGrew, 2009). Due to the interdependence of the Gf-Gc and CHC models, analogous factors are present in the CHC model. Thus, the number and nature of the second-order factors in both models can still be traced back to Thurstone's primary abilities. The CHC and Gf-Gc models are distinguished at the second-order level chiefly because the CHC model has several additional factors, most notably, reading and writing ability (Grw) and domain-specific knowledge (Gkn); these two factors are instead subsumed by Gc in the Gf-Gc model (Horn & Blankson, 2005; McGrew, 2009).

P. E. Vernon, who was a contemporary of both Thurstone and Cattell, proposed the first hierarchical model of intelligence in 1950 (Vernon, 1961). In contrast with Thurstone's primary factor model, Vernon's verbal-perceptual model contained a *g* factor and only two broad second-order group factors: the *v:ed* factor subsumed first-order factors for verbal, scholastic and numerical ability, and *k:m* was formed by loadings from first-order factors of mechanical information, spatial ability, and perceptual and psychomotor abilities (Vernon, 1961, 1965). Johnson and Bouchard (2005a) found that the addition of a second-stratum Image Rotation factor significantly improved the fit of the verbal-perceptual model, and thus they proposed the Verbal-Perceptual-Image Rotation (VPR) model as an extension of Vernon's model. This return towards a more parsimonious model similar to Vernon's was also anticipated by researchers such as Undheim (1981) and Gustafsson (1984).

The third feature in Table 3.1 indicates that the broad group factors in the VPR model are characterized by the subject-matter content of the tests (Johnson & Bouchard, 2005b). In the CHC and Gf-Gc models there is instead a contrast between factors which are theorized to involve more basic process abilities (e.g. Gf, Gsm), and those which are thought to be measures of acquired knowledge (e.g. Gc, Gq) (McGrew, 2009; Horn & Blankson, 2005). For example, Carroll (1993) stated that: "the [second-order] domains appear to differ in the relative emphasis they give to process, content and manner of response" (p. 634). However, as mentioned above, this distinction between fluid and crystallized factors was not supported in previous model-comparison studies where the VPR model outperformed Gf-Gc and three-stratum models (Johnson & Bouchard, 2005a, 2005b; Johnson, Te Nijenhuis, et al., 2007). In fact, Johnson and Bouchard (2005a) found that even in Cattell's test battery designed according to Gf-Gc theory (the Comprehensive Ability Battery;

Hakistian & Catell, 1975), the verbal-perceptual distinction was better supported than the fluid-crystallized one (as assessed by model fit).

The fluid-crystallized division also leads to theoretical problems for the Gf-Gc and CHC models. In the Gf-Gc model the contrast between ability and knowledge domains is emphasized to the exclusion of a *g* factor (Horn & Blankson, 2005); however, the *g* factor has been supported in almost all factor-analytic studies where it was possible to find one (Carroll, 1993; Jensen, 1998) and the Gf and Gc factors have a correlation as high as .85, supporting an underlying *g* factor (Johnson & Bouchard, 2005b). In the CHC model, the presence of both a *g* factor and a Gf factor is problematic because of their theoretical similarity: both factors have been described as involving the ability to reason and profit from experience across many cognitive domains (Carroll, 1993; Cattell, 1987). For example, Carroll (1993) stated that: "in the main, I accept Spearman's concept of *g*, at least to the extent of accepting for serious consideration his notions about the basic process measured by *g*—the apprehension of experience… and the eduction of relations and correlates" (p. 637), while Cattell wrote of the Gf factor that it was "a single relation perceiving capacity" that could be invested in any cognitive domain (1987, p. 138, cited in Kan, Kievet, Dolan, & van der Maas, 2011). This theoretical redundancy has been pointed out by several authors, and a number of studies have found that *g* and Gf factors are statistically indistinguishable (Gustafsson, 1984, 1988, 2002; Kan et al., 2011; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Kvist & Gustafsson, 2008; Undheim & Gustafsson, 1987); however, these models also tend to generate often unacknowledged out-of-range parameter estimates, such as negative residual variances. Recently, Kan et al. (2011) also found that, in participants with equal educational backgrounds, the Gc factor was identical with verbal comprehension, which is in conflict with the theoretical interpretation of Gc in Cattell's investment theory, but consistent with the role of the verbal factor in the VPR model (Johnson & Bouchard, 2005b).

The last feature which distinguishes the VPR model from the Gf-Gc and CHC models is the number of and nature of the first-order factors. Along with Spearman and other early British intelligence researchers, Vernon (1961) criticized American investigators for accepting too many group factors in the lower orders of the intelligence hierarchy; he argued that this was due to overly lax selection criteria and because their factor-analytic methods assigned some of the *g* variance to group factors. In favor of his more limited set of broad group factors, Vernon noted that *v:ed* and *k:m* emerged in any representative battery of tests, whereas the narrower (generally, first-order) factors proposed by American psychologists were very dependent on the particular tests administered and the selectiveness of the sample (see Appendix in Vernon, 1961). This lack of certainty about narrow factors is maintained in the VPR model, in that it does not make specific predictions about which first-order factors should emerge in a given test battery, instead leaving the characters of the factors to vary according to the specific tests in the battery (Table 3.1). Vernon (1961) also offered a pragmatic argument against naming and including narrow ability factors in the structure of intelligence; he observed that often the narrow factors in intelligence test batteries did not add substantial incremental variance to the prediction of educational or occupational performance, over and above *g* and the broad group factors. This objection is not taken into account by CHC investigators, who aim to include every factor identified in intelligence research in the CHC theory/model (McGrew, 2009), regardless of whether they add significant incremental validity over higher-order factors towards predicting outcomes of interest. Proponents of the Gf-Gc model also maintain that first-orders factors should be named, and that the factors which should emerge for a given battery can be pre-specified (Horn & Blankson, 2005).

In spite of claims for its status as the leading intelligence theory (McGrew, 2009), there are still empirical reasons to doubt whether the CHC model provides an accurate picture of the overall structure of intelligence. Carroll’s (1993) three-stratum theory was based on his interpretation of numerous exploratory factor analyses, not on confirmatory factor analysis, which allows the researcher to investigate and control many more aspects of the measurement model, and, especially, to pit competing models against each other empirically. Second, the vast majority of the datasets re-analyzed by Carroll were not suited to determining the broad higher-order structure of ability: the CHC model contains at least ten second-order factors, but all except two of Carroll’s 461 datasets contained three or fewer second-order factors.

Although the broad CHC factors have been supported in a number of more recent studies (see McGrew, 2005, for summary), much of this research has been performed on test batteries designed within the CHC framework or its precursors, and competing models have not been compared with it. These criticisms also apply to the factor-analytic evidence supporting the Extended Gf-Gc model (Horn & Blankson, 2005).

In order to establish whether the CHC, Extended Gf-Gc, or VPR model is the best-supported, further confirmatory studies are needed which compare their predictions in test batteries that were not constructed according to any particular theory of intelligence. As mentioned above, previous studies have not examined the most recent versions of these models, thus this was the main purpose for the current study.

3.1.1 Previous factor-analytic research on Project TALENT

The current study was undertaken with data from Project TALENT, which was a longitudinal study on American high school students that was designed to investigate their aptitudes, interests, and backgrounds, and the influences of these variables on educational and occupational outcomes (PT; Flanagan et al., 1962). During Project TALENT, 60 aptitude and achievement tests were given to a very large and nationally-representative sample of the U.S. student population (see Methods below for more details). As detailed below, three PT datasets were also analyzed by Carroll (1993), which provided a basis for the development of the CHC model in our study. The data are thus of particular relevance to the question of which of the three models provides the best fit. In order to provide context for the factor analysis of this test battery in this study we first review notable previous analyses of these data.

The first report which contained an analysis of the aptitude and achievement tests in PT was written by the research group who designed the study (Flanagan et al., 1964). Instead of performing factor analysis, however, Flanagan et al. (1964) examined correlation matrices and uniqueness coefficients of the tests. Using this method they tentatively identified and labeled seven common factors: general verbal ability, reasoning, rote memorization, spatial visualization, visual perception, speed of response, and information in the mechanical-electric-electronic domain.

Lohnes (1966) ran principal components analysis on a combined sample of grades 9 and 12 participants from Project TALENT. He used all 60 test scores in the battery and extracted eleven factors, excluding factors for grade and sex. However, four of these factors were highly specific (such as information on etiquette, or hunting and fishing); subsequent investigators have typically excluded a number of these information tests because of their highly specific nature, but also because they contained a small number of items and many had low reliability coefficients (see Flanagan et al., 1964).

Shaycroft (1967) examined the changes in 47 PT test scores from grade 9 to grade 12, and performed principal-axes factor analysis on the tests. She retained seven broad factors, as Lohnes (1966) did. However, the most extensive factor analysis of PT tests in this period was performed by Cureton (1968), who provided detailed comparisons of his factors to those in Lohnes (1966) and Shaycroft (1967).

Cureton's (1968) sample consisted of 543 students from Project TALENT who also completed three other intelligence test batteries. He performed three different factor analyses: on all the tests combined, the non-Talent tests only, and the PT tests only. For the PT test analysis Cureton (1968) excluded all the information tests with less than nine items, and ran principal-axes factor analysis with oblique rotation. He accepted seven factors, and although these differed slightly by sex, each model included a factor which combined the English and Math tests, a verbal-information factor, a clerical-perceptual factor, as well as factors for spatial reasoning, mechanical/outdoor knowledge, math and memory. These factors were generally consistent with those in Lohnes (1966) and Shaycroft (1967), despite different factoring methods and selection of tests in the three studies. Importantly, Cureton (1968) also observed that the mechanical factor had a tendency to combine with the spatial factor, and that the verbal-information factor was closely related to the English and Math factor; thus Cureton (1968) observed that "though second-order and hierarchical analysis was not used, the results are in striking accord with the theory of cognitive abilities outlined by Vernon" (p. 71).

When Carroll (1993) re-analyzed the data from Project Talent, he revisited the analyses of Flanagan et al. (1964), Shaycroft (1967) and Cureton (1968), which is a reflection of the lack of intervening factorial research after these seminal studies.[3] Carroll (1993) accepted seven first-order factors in his re-analysis of the grade 9 data from Shaycroft (1967), and his factors were very similar to those found by Shaycroft and Cureton, except that he did not find the Math and English tests to combine to form a factor. Despite the fact that these seven first-order factors were underneath four different second-order broad factors according to the three-stratum model, Carroll (1993) obtained only one second-order factor for the male data, which he classified as Gc (see dataset codename SHAY01; Carroll, 1993). In the female data (SHAY02), Carroll extracted three second-order factors: 2H, 2V (broad visualization) and a technical knowledge factor; he also found a third-order *g* factor. Carroll's (1993) analysis of Flanagan et al. (1964) was based on a correlation matrix which did not include the information tests (FLAN01). He extracted five first-order factors from these tests: verbal ability (V), math knowledge (KM), English language usage (EU), visualization (VZ) and perceptual speed (P). According to the three-stratum model, these factors should have loaded onto three separate higher-order factors, but Carroll (1993) only obtained one factor, which he characterized as 2H (a combination of fluid and crystallized intelligence). Together, the re-analyses by Carroll suggest that the higher-order structure of the PT tests is more parsimonious than implied by three-stratum theory, and thus potentially Gf-Gc and CHC theory as well.

Three more recent studies using PT data took Carroll's (1993) factor solutions as a starting point (Reeve, 2004; Reeve & Heggestad, 2004; Reeve et al., 2006). Although Reeve and colleagues found acceptable fit for their confirmatory factor analysis (CFA) measurement models, these studies were not primarily aimed at investigating the structure of the PT test battery, and contained no exploratory factor analysis (EFA) to determine the number of first-order factors for the selected tests, nor higher-order factor analysis.

[3] Caroll's (1993) re-analysis of Cureton (1968) was based on the dataset with the PT tests combined with three additional test batteries, and hence is not as relevant to the current review as Carroll's two other re-analyses.

In the present study we first performed EFA in order to establish the first-order structure of the PT tests. This provided an objective basis upon which to perform higher-order CFA to test the relative fits of the CHC, Extended Gf-Gc and VPR models.

3.2 Methods

3.2.1 Sample

The participants in Project TALENT (PT) were drawn from a stratified random sample of all public and private high schools in the United States in 1960 (Flanagan et al., 1962). The full obtained sample consisted of 376,213 students, with approximately 100, 000 students in each grade from 9 through 12. Of the full sample, 50.13% was female. The age range was from a mean of 14.4 in grade 9 (SD = .78) to 17.3 in grade 12 (SD = .67). The full individual age range was from 8 to 21.

3.2.2 Measures

Short descriptions and reliabilities of each cognitive ability test used in the current study are presented in Table 3.2. In addition to aptitude tests, the designers of PT included a large number of multiple-choice information tests because they sought to use these to predict future educational and vocational success in a wide variety of areas (Flanagan et al., 1962). The information tests were based partly on knowledge acquired from formal education, but were also designed to assess self-motivated learning outside the classroom (Flanagan et al., 1962). The information tests were also designed to be non-redundant with the achievement tests; thus the math information test contained factual items on mathematical concepts, but did not require problem solving as did the arithmetic and math achievement tests (Flanagan et al., 1962).

Following previous analysts such as Cureton (1968), we excluded tests with less than eight items because of their low reliabilities and tendency to form highly specific factors. An effort was also made to exclude information tests that were likely to be sex-biased due to unequal learning opportunities for boys and girls, such as the Sports, Farming and Home Economics tests. Nonetheless, the Aeronautics and

Space, Electricity and Electronics and Mechanics tests were retained because of their importance in distinguishing the VPR from the CHC and Gf-Gc models; the perceptual factor in the VPR model is formed from a combination of spatial and mechanical-knowledge factors, whereas these factors load onto separate second-stratum factors in the CHC and Gf-Gc models (Johnson & Bouchard, 2005b; Horn & Blankson, 2005; McGrew, 2009). The advanced mathematics test (R333) [4] was excluded because it included material that was not taught until grades 10 through 12 (Wise et al., 1979); thus it was deemed to be an unfair test for grade 9 students.

The remaining PT battery still contained 16 information tests that we considered were possibly less relevant to cognitive ability than the aptitude tests. Flanagan et al. (1962) defended the information tests as indicators of general intelligence based on their inclusion in classic intelligence test batteries such as the Army Alpha test and the Otis Mental Ability Tests, but they also noted that information tests are measures of interest and past achievement in specific areas (such as Biology, Physics, Literature, etc.). To defuse this question about the appropriateness of including the information tests, we fit our models to two selections of tests (hereafter termed the *broad* and *narrow* selections). The broad selection included the information tests, and consisted of 37 tests in total. The narrow selection excluded the information tests except for Vocabulary, and consisted of 22 tests.

[4] Variable ID numbers that were assigned by Project TALENT are occasionally referenced, in order to clarify which tests are being discussed.

Table 3.2
Project Talent test names, short descriptions, and reliabilities for males/females.

Test Name	Description	Items	Reliability
Vocabulary	General vocabulary questions.	21	.71/.71
Literature	Items on a broad selection of literary works.	24	.72/.70
Music	Musical information (not requiring formal training in music).	13	.67/.67
Social Studies	Items on facts and concepts from history, economics, civics, geography and current affairs.	24	.83/.79
Mathematics	Items on mathematical information and concepts.	23	.81/.78
Physical Science	Items about chemistry, physics, astronomy, and other physical sciences, not necessarily acquired through formal education.	18	.77/.72
Biological Science	Questions about botany, zoology and microbiology.	11	.57/.51
Aeronautics and Space	Items on flying technique, navigation, jet planes, and space exploration	10	.63/.34
Electricity and Electronics	Items on the construction and maintenance of electrical or electronic equipment.	20	.76/.43
Mechanics	Information on automobiles, common machines, etc.	19	.66/.48
Art	General knowledge about art, artists and art works.	12	..64/.65 [a]
Law	General knowledge items that could be acquired through books or news reports on legal affairs.	9	.51/.43 [a]
Health	Items on practical health maintenance, nutrition and common health care techniques.	9	.58/.55 [a]
Bible	General knowledge about the characters and teachings of the Bible.	15	.74/.73 [a]
Theatre and Ballet	General terms from theatre and ballet.	8	.55/.59 [a]
Miscellaneous	Miscellaneous knowledge questions.	10	.48/.42[a]
Memory for sentences	Recalling a missing word from a memorized sentence.	16	.62/.63[b]
Memory for words	Recalling an English word that corresponds to a word in a (fictional) foreign language.	24	.80/.83[b]
Disguised words	The ability to use phonetic sound to puzzle out which familiar English word a nonsense word corresponds to.	30	.86/.87[c]
Spelling	Items testing the ability to spell, and not the size of vocabulary.	16	.60/.56
Capitalization	Items requiring the correct capitalization of words in a sentence.	33	.85/.83
Punctuation	Items on the appropriate use of punctuation.	27	.72/.73
English usage	Knowledge of preferred phrasing in English.	25	.56/.49
Effective expression	Items testing the ability to recognize whether an idea has been expressed clearly, concisely and smoothly.	12	.63/.52

Word functions in sentences	A test of sensitivity to grammatical structure. The test taker must find the word that performs the same grammatical function as a word in another sentence.	24	.81/.84
Reading comprehension	Multiple-choice items on the comprehension of a written passage.	48	.86/.84[c]
Creativity	Verbal items requiring ingenious solutions to practical problems.	20	.73/.68
Mechanical reasoning	Items of the ability visualize the operation of physical force, such as the effect of gravitation, gears, pulleys, levers, etc.	20	.76/.64
Visualization in 2 dimensions	Items requiring mental rotation of shapes in two dimensions.	24	.81/.80[c]
Visualization in 3 dimensions	Items on the ability to visualize a how a two-dimensional figure would look after it were folded into a three-dimensional one.	16	.70/.59
Abstract reasoning	A non-verbal test on the ability to identify the logical progression of elements in a complex pattern (similar to Raven's progressive matrices).	15	.66/.65
Math 1 Arithmetic reasoning	A test of the ability to reason in the manner required to solve arithmetic problems, with only very simple computation.	16	.73/.71
Math 2 Introductory High School mathematics	A test of mathematics taught up to an including the 9th grade, including items on algebra, fractions, simple geometry, etc.	24	.78/.73
Arithmetic computation	A test of the speed and accuracy of addition, subtraction, multiplication and division.	72	Not avail.
Table reading	A test on the speed and accuracy of obtaining information from a table.	72	Not avail.
Clerical checking	A test on the speed and accuracy of checking whether two pairs of names are identical.	74	Not avail.
Object inspection	A test on the speed and accuracy of spotting small differences between objects when comparing them visually.	40	Not avail.

Note: Descriptions adapted from Wise et al. (1979). Reliability estimates taken from Flanagan et al. (1964, Table 2-5), and are based on the Kuder-Richardson Formula 21 (Kuder & Richardson, 1937) unless otherwise noted. The reliabilities are lower-bound estimates, and are based on the mean for all grades combined.

[a] Estimate based on Kuder-Richardson Formula 20.

[b] Estimate may be an overestimate due to lack of experimental independence of items (Flanagan et al., 1964).

[c] Split-half reliability estimate.

3.2.3 Data preparation

Among the PT tests was a screening test consisting of basic knowledge questions that were taught in elementary school; it was designed to identify students who were functionally illiterate, mentally retarded, or who displayed an apathetic attitude to the tests (Wise et al., 1979). A response credibility index was available for all participants according to their scores on the screening test, taking into account whether the score could be explained by illiteracy (a low score on the reading comprehension test—R250), mental slowness (a low score on the clerical checking test—A430), problems with clerical inaccuracy (a low percentage correct on the clerical checking test—P430), or some combination of these. We removed cases who scored below the threshold for the screening test, except those cases for which no explanation was provided by their scores on the other three tests. Also, in order not to restrict the range of cognitive ability, participants who failed the screening test ostensibly due to mental slowness were left in the sample. Students with missing scores on the screening test were also retained.

Prior to the analysis, data were screened for normality and outliers. Three tests were found to have problematic violations of normality in each grade: Capitalization and English Usage were found to be negatively skewed, and Table Reading was positively skewed (all three also displayed leptokurtic distributions). To deal with these violations, a logarithmic transformation was applied to Capitalization, a square-root transformation to English Usage, and logarithmic and cosine transformations were applied to Table Reading. The scores for Capitalization and English usage were reflected prior to transformation, and re-reflected afterwards in order to keep the original direction of the scores. The scores for Table Reading were also re-reflected after transformation because the cosine transformation reflected them. Following transformation, within each grade the highest remaining skewness was for Clerical Checking (z = 0.70-1.08) and the highest remaining kurtosis was for Table Reading (z = 1.74-2.63).

After transformation, no extreme univariate outliers remained given the large sample size. In order to control for potential multivariate outliers the Mahalanobis distance and Cook’s distance were obtained for complete cases (separately in the broad and narrow selections). These statistics were obtained by regressing the PT student ID number (a random variable) onto the test scores. Cases that had a Mahalanobis distance with a $p < .001$ ($\chi^2(37) = 69.35$, for the broad selection, $\chi^2(22)$

= 48.27, for the narrow selection), and a Cook's distance of greater than 4/N were removed from the sample (critical values suggested by Tabachnick & Fidell, 2007). Following data screening, total sample size was reduced to 366,857 in the broad selection, and 366,695 in the narrow selection (2.49 - 2.54 % of the sample removed).

We handled missing data by using multiple imputation with five datasets for the exploratory analyses and direct maximum likelihood estimation for the confirmatory factor analyses. These missing-data methods yield the same results (Brown, 2006), and require the assumption that the data be missing at random (MAR). For this assumption to have been violated, students would have had selectively to avoid particular tests specifically due to awareness of lower ability in those areas. Given that 2.27-3.23% of all test scores were missing across each grade, usually in relatively large 'clumps' for individuals, this is unlikely to have occurred enough to affect the results. A comparison of means and correlations in the full dataset to those with listwise deletion also showed only very small differences, indicating that the pattern of results would be the same basically no matter how missing data were treated.

3.2.4 Analysis method

Despite the existence of previous such analyses in PT, we used exploratory factor analysis to estimate the factor structure given our particular selections of tests and data screening methods (for example, our treatment of outliers and missing data). Consistent with previous analyses of the PT data, factor analyses were performed separately for each combination of grade and sex (Carroll, 1993; Reeve et al., 2006). This was done to retain a number of replication samples and to identify possible differences in the factor structures across the sexes and grades. In order to determine the numbers of factors to extract in the exploratory analyses, parallel analysis and Velicer's minimum average partial (MAP) test were obtained for each sample in SPSS (see O'Connor, 2000, for syntaxes); the Kaiser criterion (the number of eigenvalues > 1) was also examined. Most important, however, was a consideration of the interpretability of the factors and whether each factor had at least two tests whose highest loadings were there (a criterion suggested by Carroll, 1993). All analyses were performed with maximum likelihood estimation, and Promax rotation (Kappa set to 4) was used for all the exploratory analyses.

For the CFA results, we report three conventional fit indices: the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI) and Standardized Root Mean Square Residual (SRMR) (Brown, 2006). An information criterion index, the saBIC (sample size-adjusted BIC), was selected in order to be able to directly compare our non-nested models. The saBIC was chosen because the unadjusted BIC imposes a high penalty for additional parameters based upon sample size (Kenny, 2011); the saBIC has also been found to perform well in model selection (Sclove, 1987; Yang, 2006). Although not shown, AIC (Akaike Information Criterion) statistics gave rise to the same conclusions as the saBIC.

3.3 Results

3.3.1 Exploratory factor analysis

3.3.1.1 Broad selection

For the males in the broad selection, parallel analysis and the MAP test each indicated that there were four factors, while the Kaiser criterion suggested five factors (except in grade 9 males, where it suggested four). However, when a sixth factor was extracted it was clearly interpretable as the math knowledge factor that was identified by previous researchers (Cureton, 1968; Carroll, 1993); the factor had the highest loadings for Math Information and the second part of the Math test. In contrast, a seventh factor was not clearly interpretable and contained no highest loading from any test. Thus, we retained six factors in each male sample.

In the female samples, parallel analysis suggested there were only three factors, but the MAP test indicated four factors in grades 9 and 10, and five in grades 11 and 12. The Kaiser criterion suggested five factors throughout. Nonetheless, the sixth factor was identified as the Math factor in the same manner as in the male samples. In grades 9 and 10 the seventh factors were singlets with the highest loadings for Disguised Words and Memory for Sentences, respectively. In grades 11 and 12, the seventh factors had two highest loadings (Memory for Sentences and Bible in grade 11, and Biological Sciences and Bible in grade 12), but these factors, unlike the others, varied in each sample and were difficult to interpret; thus we retained six factors in each female sample as well.

Table 3.3 presents the factor pattern matrices for grade 10 males and females, with loadings under .15 suppressed (.15 was used as a cutoff to determine whether a factor loading was substantively meaningful). Table 3.5 contains the factor correlation matrices. The six factors were labelled in a manner similar to Cureton (1968): the Information factor was formed largely by loadings from the information tests. The English/Math factor had loadings from the English and Math achievement tests. The Spatial/Reasoning factor was formed by tests with visual-spatial content, but also tests that involved a reasoning component (such as Creativity, Math 1). The Mechanical/Science factor had loadings from tests requiring knowledge in mechanics, electronics and science subjects. The Speed factor was formed by loadings from all the speeded tests in the battery. Finally, the Math factor had its highest loadings from tests requiring math knowledge (as opposed to tests of computation skills, which loaded more highly upon the English/Math factor).

Table 3.3
Factor pattern matrices for grade 10 males/females in the broad selection of PT tests.

Test Name	Factor					
	Inform-ation	English/ Math	Spatial/ Reasoning	Mechanical/ Science	Speed	Math
Vocabulary	.55/.56	–/.16		.28/.20		
Literature	.93/.84					
Music	.75/.74					
Social Studies	.73/.62					
Mathematics	.26/.21					.54/.64
Physical Science	.35/–			.40/.52		.24/.16
Biological Science	.41/.23			.31/.40		
Aeronautics and Space	.48/.33			.37/.33		
Electronics				.74/.68		
Mechanics			–/.18	.75/.55		
Art	.89/.90					
Law	.60/.55					
Health	.53/.45	.17/.29		.16/.17		
Bible	.69/.49					
Theatre and Ballet	.78/.88					
Miscellaneous	.65/.57					
Memory for sentences		.29/.36				
Memory for words	.17/.15	.34/.38				
Disguised words	.33/.37	.42/.43			.24/.25	
Spelling		.72/.81				
Capitalization		.63/.71				
Punctuation		.72/.76				
English usage		.62/.62				
Effective expression	–/.15	.57/.54				
Word functions in sent.		.42/.40	.16/.15			.31/.33
Reading comprehension	.58/.54	.33/.33	.15/.15			
Creativity	.28/.29	.19/–	.29/.32	.15/–		
Mechanical reasoning			.62/.69	.36/.15		
Visualization in 2D			.58/.59		.20/.18	
Visualization in 3D			.78/.77			
Abstract reasoning		.17/.25	.58/.57			
Math 1		.38/.35	.18/.23			.27/.24
Math 2		.32/.26				.59/.62
Arithmetic comp.		.52/.63			.30/.26	.26/.17
Table reading		–/.17			.66/.64	
Clerical checking					.72/.72	
Object inspection			.30/.27		.62/.63	

As seen in Table 3.3, the general pattern of loadings was highly similar across the sexes, with some minor exceptions. The loadings for physical science, biology and aeronautics tests on the Information factor were lower in females, which may be attributable to their lower reliabilities in females (see Table 3.2). However, the loadings of these tests on the Mechanical/Science factor were similar in each sex, suggesting that they functioned equally well as tests of specific mechanical/science knowledge in females as in males, but that they were better tests of general knowledge for males than females. Another interesting sex difference was that the test of mechanical knowledge loaded at .18 on the Spatial/Reasoning factor in females, but below .15 in males. This suggests that mechanical knowledge was tied more closely to spatial ability in females.

Differences in salient loadings across the eight samples were later used for the construction of the confirmatory factor models. The differences were as follows:

Relative to grade 10 males, in grade 9 males the Vocabulary test had a loading on English/Math, Reading Comprehension did not have a loading on Spatial/Reasoning, Creativity did not load substantively (had a loading below .15) onto the Mechanical/Science factor, and Punctuation had a loading on the Math factor. In grade 11 males, Memory for Words did not load substantively on the Information factor. In grade 12 males, the Vocabulary test had a loading on English/Math and the Electricity and Electronics test had a loading on the Math factor.

Relative to grade 10 males, in grade 10 females Effective Expression loaded on the Information factor, while Physical Science Information did not[5], Vocabulary and Table Reading loaded on English/Math, while Creativity did not, and Mechanical Information loaded on Spatial/Reasoning.

In grade 9 females, relative to grade 10 females, Memory for Words and Effective Expression did not load on Information, Bible had a loading on

[5] This was a logical finding given that most students were probably not exposed to formal Physics classes until higher grade levels. However, unlike in the female samples, in grade 9 and 10 males Physical Science did load on Information, perhaps because boys were more likely to be exposed to Physics knowledge outside of school.

English/Math, Word Functions in Sentences and Reading Comprehension did not load on Spatial/Reasoning, Mechanical reasoning did not load on the Mechanical/Science factor, and Social Science information loaded on the Math factor but Physical Science did not.

In grade 11 females, relative to grade 10 females, Physical Science and English Usage loaded on Information, Table Reading did not load on English/Math, and Health information did not have a loading on Mechanical/Science.

Finally, in grade 12 females, relative to grade 10 females, Physical Science and English Usage loaded on Information, while Memory for Words did not, Vocabulary and Table Reading did not load on English/Math, and Vocabulary and Health information did not load on Mechanical/Science.

3.3.1.2 Narrow selection

In males, parallel analysis and the MAP test indicated three factors. The Kaiser criterion suggested three factors in grades 9 and 10, and four factors in grades 11 and 12. In females, all the criteria suggested three factors. Nonetheless, five factors were retained for both sexes because the fourth and fifth factors corresponded to factors in the broad selection, and contained the highest loadings from at least two tests.

The sixth factor was not retained for multiple reasons. In each male sample the sixth factor was a singlet with Memory for Sentences, and thus was clearly not interpretable. In grade 9 females, there were no highest loadings on the sixth factor; in grade 10 there was a doublet with Memory for Words and Memory for Sentences, and in grade 11 and 12 females there was once again a singlet with Memory for Sentences. Although there was some evidence for a Memory factor in females, we considered that the relation between the memory tests would be best handled with correlated error variances instead of a factor, because of the similarity in the format of the tests (memorization of words or sentences, followed by multiple-choice items testing recall).

Table 3.4 displays the factor pattern matrices for grade 10 males and females, and Table 3.5 the factor correlations. Four of the factors were highly similar to those in the broad selection and were labelled the same: English/Math, Spatial/Reasoning, Speed and Math. Due to the removal of the science and mechanical information tests, there was no longer a factor for them. The fifth factor had the highest loadings for the Vocabulary and Creativity tests and was labelled the Verbal factor because all its loadings came from tests with verbal subject-matter content.

Table 3.4
Factor pattern matrices for grade 10 males/females in the narrow selection of PT tests.

Test Name	Factor				
	English/Math	Spatial/ Reasoning	Speed	Math	Verbal
Vocabulary	.36/.43				.50/.48
Memory for sentences	.18/.23				
Memory for words	.38/.38				–/.15
Disguised words	.58/.60		.26/.27		.27/.24
Spelling	.79/.81				
Capitalization	.69/.71				
Punctuation	.81/.77				
English usage	.71/.72				
Effective expression	.60/.59				
Word functions in sent.	.44/.40			.29/.30	
Reading comprehension	.51/.52				.45/.44
Creativity	.22/.20	.24/.24			.40/.37
Mechanical reasoning		.72/.66			.18/–
Visualization in 2D		.60/.60	.20/.17		
Visualization in 3D		.78/.76			
Abstract reasoning	.21/.28	.54/.53			
Math 1	.21/.24	–/.16		.43/.38	.15/.16
Math 2	.17/.23			.68/.58	
Arithmetic comp.	.28/.38		.30/.25	.44/.39	
Table reading			.66/.66		
Clerical checking			.73/.73		
Object inspection		.28/.26	.62/.63		

Relative to grade 10 males, the only factor loading differences in males were that Math 1 loaded on Spatial/Reasoning in grades 11 and 12.

Relative to grade 10 males, in grade 10 females, Math 1 had a loading on Spatial/Reasoning, Memory for words had a loading on Verbal, and Mechanical reasoning did not load on the Verbal factor.

In grade 9 females, relative to grade 10 females, Memory for Words had a loading on the Math factor (allowed only in the VPR model), and Memory for Words and Math 1 did not load on Verbal.

In grade 11 and grade 12 females, relative to grade 10 females, Math 2 did not load on English/Math, and Memory for Words and Math 1 did not load on Verbal.

Table 3.5
Factor correlation matrices for grade 10 males (below diagonal) and females (above diagonal) in the broad and narrow selections of PT tests.

	Info.	English/ Math	Spatial/ Reasoning	Math	Speed	Mech./ Science
Information	–	.783	.659	.641	.093	.722
English/Math	.760	–	.676	.688	.196	.570
Spatial/Reas.	.593	.593	–	.571	.186	.602
Math	.612	.607	.500	–	.064	.614
Speed	.099	.201	.140	.086	–	.008
Mech./Science	.692	.482	.642	.504	-.037	–
Narrow Selection						
Verbal	–	.582	.595	.513	.070	
English/Math	.590	–	.626	.694	.267	
Spatial/Reas.	.597	.561	–	.593	.294	
Math	.494	.737	.577	–	.175	
Speed	.068	.211	.163	.176	–	

3.3.2 Confirmatory factor analyses

3.3.2.1 Broad selection

Based upon the results of the exploratory analysis, we developed confirmatory models for the VPR, CHC and Gf-Gc models (see Figures 3.1-3.3). Table 3.6 displays the loadings of the tests on the first-order factors for the grade 10 male sample. Differences in factor loadings compared to this sample were noted above in section 3.1.1. Although the number of factor loadings varies across the models, the number of input variables was the same for each. Supplemental tables A1 and A2 provide the numerical first-order loadings for the grade 10 males and females for all models (see Appendix A).

In order to represent the three models under consideration accurately, factor loadings were only included if they were consistent with the theories behind the three models, taking into account whether test content could account for any indicated cross-loadings. For the Gf-Gc and CHC models, for example, there was no theoretical or content rationale for the math achievement and arithmetic tests to load onto a factor otherwise dominated by English tests (these factors were characterized as English achievement in the CHC model (following Carroll, 1993)[6], and Verbal comprehension in the Gf-Gc model). A hybrid first-order English/Math factor is also not specified in the models, and the first-order factors for quantitative ability and English achievement are theorized to load onto separate second-order factors in both (Horn & Blankson, 2005; McGrew, 2005). In contrast, a factor combining English and Math tests is theoretically plausible within the VPR model because factors formed by these tests are in the same domain, underneath the broad Verbal (formerly verbal-educational) factor (Johnson & Bouchard, 2005b), and because the VPR model does not pre-specify factor content. Additionally, three verbal tests were found to load onto the Math factor (Word Functions in Sentences, Memory for Words, and Punctuation). These cross-loadings were included in the VPR model because of the theoretically-based connection between the Math and English factors, but were not allowed in the CHC and Gf-Gc models.

[6] See dataset SHAY01 in Carroll (1993). This factor was also characterized as English language usage in the other PT datasets, but the label of English achievement seemed more appropriate to us.

Table 3.6
First-order loadings for the CHC, Extended Gf-Gc and VPR models in the broad selection (grade 10 males).

Test	Primary loadings			Secondary loading(s)		
	CHC[a]	**Gf-Gc**[b]	**VPR**	**CHC**	**Gf-Gc**	**VPR**
Vocabulary	K0	Vi	Information	K1	Science	Mechanical/Science
Literature	K0	Vi	Information			
Music	K0	Vi	Information			
Social Studies	K0	Vi	Information			
Mathematics	KM	Gq	Math	K0	Vi	Information
Physical Science	K1	Science	Mechanical/Science	K0, KM	Vi, Gq	Information, Math
Biological Science	K0	Vi	Information	K1	Science	Mechanical/Science
Aeronautics and Space	K0	Vi	Information	K1	Science	Mechanical/Science
Electronics	K1	Science	Mechanical/Science			
Mechanics	K1	Science	Mechanical/Science			
Art	K0	Vi	Information			
Law	K0	Vi	Information			
Health	K0	Vi	Information	A6, K1	V, Science	English/Math, Mechanical/Science
Bible	K0	Vi	Information			
Theatre and Ballet	K0	Vi	Information			
Miscellaneous	K0	Vi	Information			
Memory for sentences	A6	V	English/Math			
Memory for words	A6	V	English/Math	K0	Vi	Information
Disguised words	A6	V	English/Math	K0, P	Vi, P	Information, Speed
Spelling	A6	V	English/Math			
Capitalization	A6	V	English/Math			
Punctuation	A6	V	English/Math			
English usage	A6	V	English/Math			
Effective expression	A6	V	English/Math			
Word functions in sent.	A6	V	English/Math	Vz	Visualization	Math, Spatial/Reasoning
Reading comprehension	K0	Vi	Information	A6, Vz	V, Visualization	English/Math, Spatial/Reasoning
Creativity	Vz	Visualization	Spatial/Reasoning	K0, A6, K1	Vi, V, Science	Information, English/Math,

						Mechanical/Science
Mechanical reasoning	Vz	Visualization	Spatial/Reasoning	K1	Science	Mechanical/Science
Visualization in 2D	Vz	Visualization	Spatial/Reasoning	P	P	Speed
Visualization in 3D	Vz	Visualization	Spatial/Reasoning			
Abstract reasoning	Vz	Visualization	Spatial/Reasoning	A6	V	English/Math
Math 1	KM	Gq	English/Math	Vz	Visualization	Math, Spatial/Reasoning
Math 2	KM	Gq	Math	A6	V	English/Math
Arithmetic comp.	KM	Gq	English/Math	P	P	Speed, Math
Table reading	P	P	Speed			
Clerical checking	P	P	Speed			
Object inspection	P	P	Speed	Vz	Visualization	Spatial/Reasoning

[a] K0 = general verbal information, K1 = science knowledge, A6 = English achievement, KM = math achievement, P = perceptual speed. [b] Vi = general information, V = verbal comprehension, Gq = quantitative knowledge.

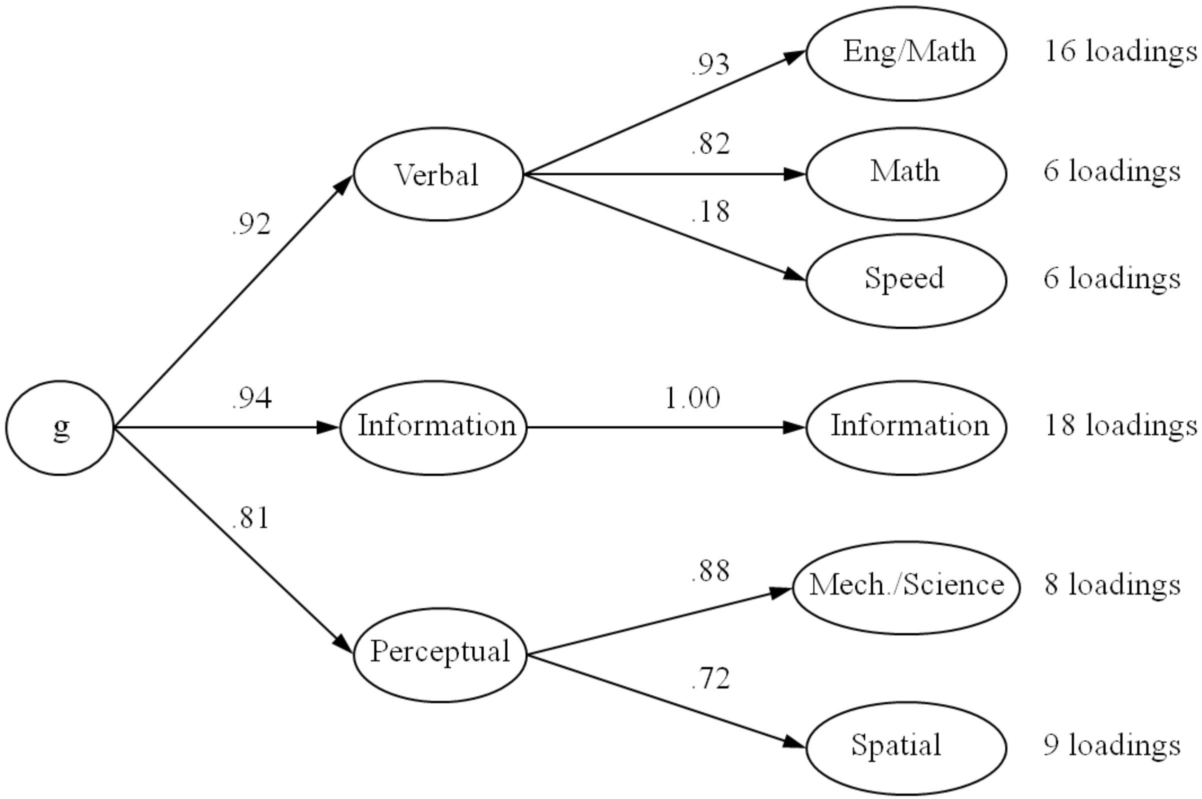

Fig. 3.1. Measurement model of the VPR model with factor loadings from the grade 10 male sample.

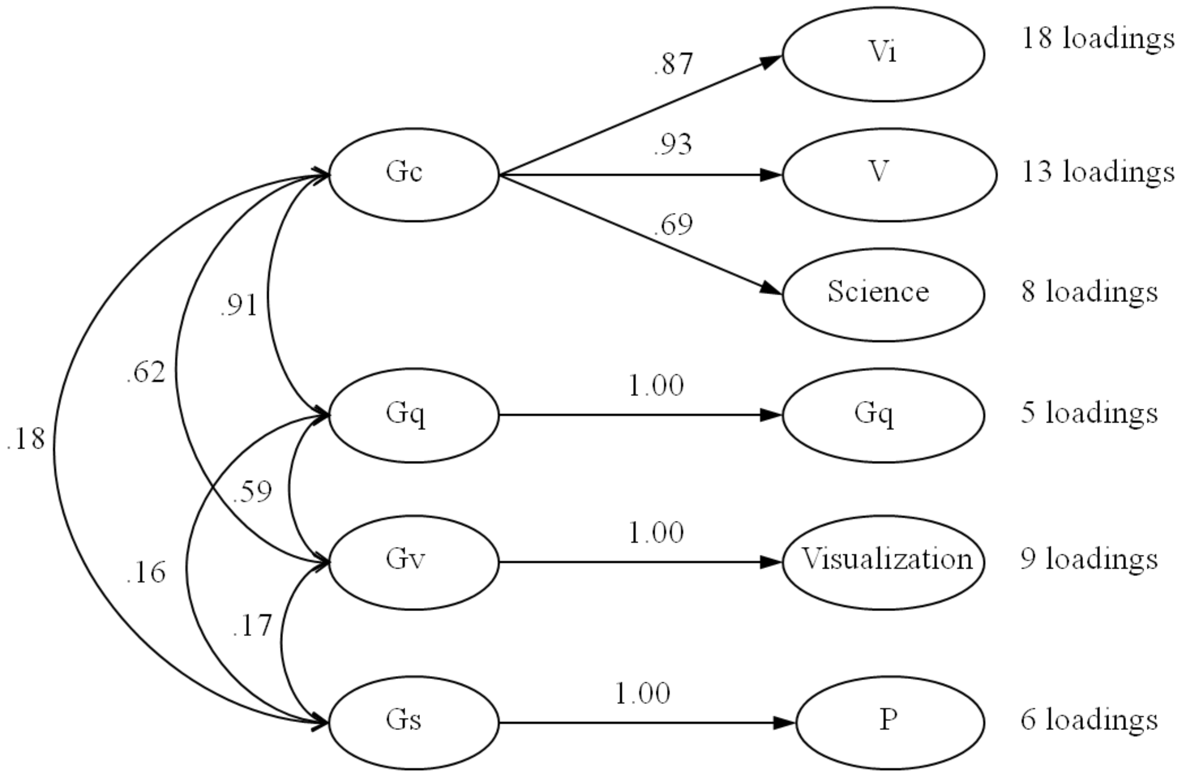

Fig. 3.2. Measurement model of the Extended Fluid-Crystallized model with factor loadings from the grade 10 male sample.

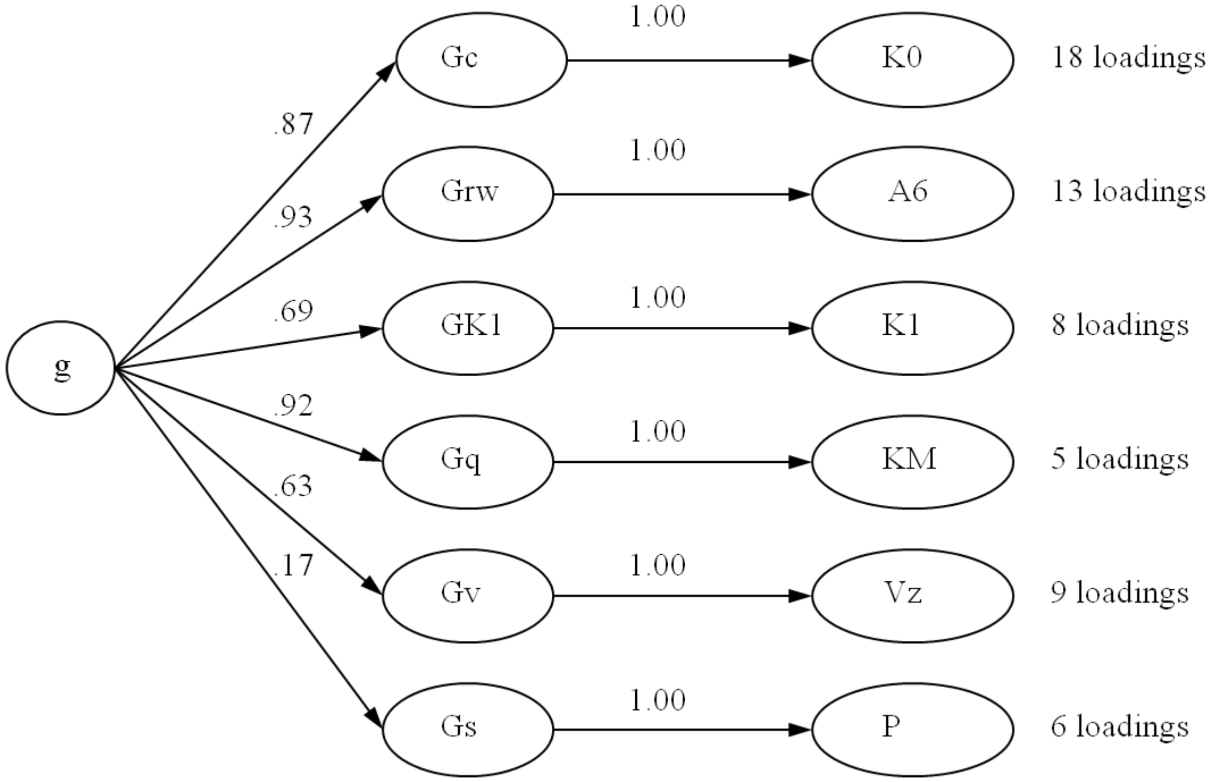

Fig. 3.3. Measurement model of the Cattell-Horn-Carroll model with factor loadings from the grade 10 male sample.

According to the CHC framework, the six first-order factors that we obtained each belonged underneath separate second-order factors (McGrew, 2009). Unfortunately, these second-order factors could not be identified with only one loading, thus the CHC model here effectively consisted only of first-order factors loading upon a second-order *g* factor. The names of the second-order factors are included in Figure 3.3 only for illustration purposes as they contributed nothing to the estimation of the model.

In the course of fitting the VPR model, a negative residual was encountered for the second-order Verbal factor, which was formed from the first-order factors Information, English/Math and Math (in males $z = -9.08$ to -5.92, $p < .001$; in females this residual was either negative or non-significant, $z = -1.48$ to 5.80). This suggested that too much of the test variance was being assigned to the Verbal factor, possibly because of the large number of tests forming the factors composing it (see Figure 3.1). In order to resolve this issue, the Information factor was placed on its own second-order factor (also named the Information factor), which eliminated the negative residual and improved model fit. Based on the exploratory analysis it was apparent that the first-order Speed factor, which consisted of mainly clerical-type speed tests, might load onto the Verbal factor formed by the Math and English tests; the first-order Speed factor's highest correlation was with the English/Math factor (see Table 3.5). Vernon (1961) observed that factors formed by clerical-type tests often loaded within the verbal-educational domain. As predicted, placing the Speed factor on the Verbal factor improved the fit of the model compared to the Speed factor having a separate loading on *g*, and this model was used as the final version of the VPR model.

The VPR, CHC and Gf-Gc models fit well in both males and females samples according to conventional fit criteria (Hu & Bentler, 1999): the RMSEA was below .060 in all cases, the SRMR below .080, and the CFI was close to .950 or greater (Table 3.7). The VPR model demonstrated the best fit in all the samples according to all fit indices that were used, followed primarily by the CHC model. A BIC difference of 10 is normally considered very strong evidence in favour of the model with the smaller value (A. E. Raftery, 1995); however, because of the large PT sample sizes the saBIC values for our models were so large that a difference of 10 was trivial. Therefore, we calculated what the saBIC differences would have been if the samples were a more conventional 500.[7] As shown in Table 3.7, the VPR model had the lowest saBIC at a sample size of 500 compared with the CHC and Gf-Gc models in each sample. In males, the VPR had a saBIC 55.0 – 93.6 lower than the CHC model, and 61.9 – 100.3 lower than the Gf-Gc model. In females, the VPR had a saBIC 36.9 – 119.3 lower than the CHC model, and 43.4 – 115.10 lower than the Gf-Gc model. The VPR model had the lowest saBIC despite being the least parsimonious model (it containined the highest number of freely-estimated parameters). Thus, the VPR model was found to consistently have the best fit to the broad selection of PT tests. The CHC model had a lower saBIC than the Gf-Gc model in five out of eight samples, but the saBIC difference was lower than 10 in all cases, thus in general the difference in fit between these models was marginal.

[7] The formula for the saBIC is -2(log-likelihood) + $p$ln[(N +2)/24], where p is the number of freely-estimated parameters. Since log-likelihood is linearly related to sample size, it was scaled by the ratio of 500 to the full sample size for each sample. In the second half of the equation, 500 was entered for N.

Table 3.7
Fit statistics of confirmatory factor models for the broad selection of PT tests.

Sample	Sample size	χ^2	*df*	saBIC	RMSEA (95% CI)	CFI	SRMR	saBIC, sample of 500
				Males				
VPR model								
Grade 9	49264	45592.64	592	8755769.82	.039 (.039-.040)	.955	.036	89301.38
Grade 10	48561	51421.27	593	8761767.73	.042 (.042-.042)	.951	.037	90652.44
Grade 11	44172	51480.47	594	8015636.24	.044 (.044-.044)	.949	.036	91163.59
Grade 12	38894	47517.10	592	7040873.99	.045 (.045-.045)	.948	.036	90949.57
Cattell-Horn-Carroll Model								
Grade 9	49264	52812.82	599	8762936.61	.042 (.042-.042)	.948	.039	89356.34
Grade 10	48561	59600.54	599	8769901.32	.045 (.045-.045)	.943	.040	90715.45
Grade 11	44172	61253.79	600	8025364.45	.048 (.048-.048)	.939	.040	91255.98
Grade 12	38894	56212.83	598	7049525.37	.049 (.049-.049)	.938	.041	91043.11
Fluid-Crystallized model								
Grade 9	49264	52597.02	596	8762743.69	.042 (.042-.042)	.948	.038	89363.28
Grade 10	48561	59357.72	596	8769681.34	.045 (.045-.045)	.943	.040	90722.07
Grade 11	44172	60878.92	597	8025012.13	.048 (.047-.048)	.939	.040	91263.89
Grade 12	38894	55655.29	595	7048990.01	.049 (.048-.049)	.939	.040	91045.06
				Females				
VPR model								
Grade 9	49973	36559.98	595	8863188.52	.035 (.034-.035)	.962	.025	90385.69
Grade 10	48237	37640.05	591	8598498.30	.036 (.036-.036)	.961	.025	88974.32
Grade 11	46504	40765.08	591	8329464.47	.038 (.038-.039)	.958	.032	94724.80
Grade 12	41119	42975.16	594	7378914.32	.042 (.041-.042)	.951	.042	95289.34
Cattell-Horn-Carroll Model								
Grade 9	49973	41993.83	601	8868576.52	.037 (.037-.037)	.956	.026	90422.60
Grade 10	48237	44907.56	597	8605720.18	.039 (.039-.040)	.954	.027	89030.91
Grade 11	46504	50429.74	597	8339083.71	.042 (.042-.043)	.948	.034	94815.95
Grade 12	41119	53673.68	600	7389568.16	.046 (.046-.047)	.938	.037	95408.62
Fluid-Crystallized model								
Grade 9	49973	41739.25	598	8868344.86	.037 (.037-.037)	.956	.026	90429.14
Grade 10	48237	44481.76	594	8605317.19	.039 (.039-.039)	.954	.026	89035.65
Grade 11	46504	49579.24	594	8338255.94	.042 (.042-.042)	.949	.033	94815.45
Grade 12	41119	52638.67	597	7388555.49	.046 (.046-.046)	.939	.036	95404.44

3.3.2.2 Narrow selection

Confirmatory factor models for the narrow selection of tests were developed in the same manner as those for the broad selection, based upon the exploratory analysis (see Table 3.8 for the first-order loadings; differences in the loadings across each sample were detailed in 3.1.2.). Once again cross-loadings for Math tests onto the English achievement (in the CHC model) and Verbal comprehension factor (in the Gf-Gc model) were fixed to zero, as was the loading of Word Functions in Sentences on their Math factors.

The initial VPR model fit properly in all grades except in grade 9 males, where a small negative residual was encountered for the Math factor (standardized loading = -.034). Fixing this residual to zero resulted in another negative residual for the Math 2 test, and a negative loading for this test on the English/Math factor (standardized loading = - .35), indicating that this factor loading was the source of model misfit. Thus, the Math 2 loading was removed from English/Math, which resolved the negative residual.

The second-order structures of the models were the same as in the broad selection, except the absence of a first-order Mechanical/Science factor meant that there was no longer a second loading onto the Perceptual factor in the VPR model or a third loading on the Gc factor in the Gf-Gc model. In the CHC model, the second-order factor for science knowledge (K1) was absent. The Verbal factor occupied the same role as the Information factor did in the broad selection.

Table 3.9 contains the fit statistics for the models based on the narrow selection. The VPR model again fit the data best according to all criteria (except for two comparisons where the SRMRs were equal). The second best-fitting model was generally the CHC model, but the saBIC differences between the CHC model and Gf-Gc model were again small and inconsistent, pointing to only a marginal difference overall between them. In the hypothetical male samples of 500, the VPR had a saBIC 8.4 – 17.5 lower than the CHC model, and 14.8 – 24.6 lower than the Gf-Gc model. In females, the VPR had a saBIC 15.8 – 34.4 lower than the CHC model, and 21.3 – 30.6 lower than the Gf-Gc model. Although the saBIC difference between the VPR and CHC model was less than 10 in grade 9 males, a difference of 8.4 is still characterized as “strong” evidence of a model fit difference, and corresponds to a posterior odds of 66:1 in favor of the VPR model (Raftery, 1995). The AIC, RMSEA and CFI also favored the VPR over the CHC model. Thus, the VPR model demonstrated the overall best fit to the narrow selection of PT tests.

Table 3.8
First-order loadings for the CHC, Extended Gf-Gc and VPR models in the narrow selection (grade 10 males).

Test	Primary loadings			Secondary loading(s)		
	CHC[a]	**Gf-Gc**[b]	**VPR**	**CHC**	**Gf-Gc**	**VPR**
Vocabulary	K0	Vi	Verbal	A6	V	English/Math
Memory for sentences	A6	V	English/Math			
Memory for words	A6	V	English/Math			
Disguised words	A6	V	English/Math	K0, P	Vi,P	Verbal, Speed
Spelling	A6	V	English/Math			
Capitalization	A6	V	English/Math			
Punctuation	A6	V	English/Math			
English usage	A6	V	English/Math			
Effective expression	A6	V	English/Math			
Word functions in sent.	A6	V	English/Math			Math
Reading comprehension	A6	V	English/Math	K0	Vi	Verbal
Creativity	K0	Vi	Verbal	Vz, A6	Visualization, V	Spatial/Reasoning, English/Math
Mechanical reasoning	Vz	Visualization	Spatial/Reasoning	K0	Vi	Verbal
Visualization in 2D	Vz	Visualization	Spatial/Reasoning			
Visualization in 3D	Vz	Visualization	Spatial/Reasoning			
Abstract reasoning	Vz	Visualization	Spatial/Reasoning	A6	V	English/Math
Math 1	KM	Gq	Math	P	P	English/Math, Speed
Math 2	KM	Gq	Math			English/Math
Arithmetic comp.	KM	Gq	Math	P	P	Speed, English/Math
Table reading	P	P	Speed			
Clerical checking	P	P	Speed			
Object inspection	P	P	Speed	Vz	Visualization	Spatial/Reasoning

[a] K0 = general verbal information, A6 = English achievement, KM = math achievement, P = perceptual speed.
[b] Vi = general information, V = verbal comprehension, Gq = quantitative knowledge.

Table 3.9
Fit statistics of confirmatory factor models for the narrow selection of PT tests.

Sample	Sample size	χ^2	*df*	saBIC	RMSEA (95% CI)	CFI	SRMR	saBIC in sample of 500
				Males				
VPR model								
Grade 9	49096	18297.50	186	5724268.66	.045(.044-.045)	.962	.041	58560.39
Grade 10	48254	18867.29	185	5743450.08	.046(.045-.046)	.962	.040	59448.13
Grade 11	44165	17562.76	184	5254414.22	.046(.046-.047)	.963	.037	59755.12
Grade 12	38877	16526.62	184	4611559.86	.048(.047-.048)	.962	.037	59577.66
Cattell-Horn-Carroll Model								
Grade 9	49096	20313.12	190	5726253.80	.046(.046-.047)	.958	.041	58568.75
Grade 10	48254	21446.22	190	5745990.95	.048(.047-.049)	.957	.041	59459.50
Grade 11	44165	20453.84	189	5257267.72	.049(.049-.050)	.957	.039	59772.65
Grade 12	38877	18794.95	189	4613791.25	.050(.050-.051)	.956	.039	59591.63
Fluid-Crystallized model								
Grade 9	49096	20045.70	187	5726009.25	.047(.046-.047)	.959	.041	58575.15
Grade 10	48254	21272.96	187	5745840.53	.048(.048-.049)	.957	.041	59466.84
Grade 11	44165	20276.00	186	5257112.43	.049(.049-.050)	.957	.038	59779.76
Grade 12	38877	18555.20	186	4613573.67	.050(.050-.051)	.957	.038	59597.67
				Females				
VPR model								
Grade 9	49922	16921.93	185	5815009.26	.043(.042-.042)	.967	.037	59487.47
Grade 10	48209	16689.62	184	5648320.43	.043(.043-.044)	.968	.036	58470.86
Grade 11	46499	17314.84	187	5468164.36	.044(.044-.045)	.965	.035	62166.16
Grade 12	41095	16920.67	187	4830928.59	.046(.046-.047)	.961	.034	62390.14
Cattell-Horn-Carroll Model								
Grade 9	49922	20269.31	191	5818310.80	.045(.045-.046)	.960	.038	59503.31
Grade 10	48209	19909.80	189	5651502.59	.047(.046-.047)	.961	.038	58488.84
Grade 11	46499	21025.72	191	5471844.96	.048(.048-.049)	.958	.037	62196.01
Grade 12	41095	21537.53	191	4834515.66	.051(.050-.052)	.953	.037	62424.49
Fluid-Crystallized model								
Grade 9	49922	19912.12	188	5817976.53	.046(.045-.046)	.961	.038	59508.80
Grade 10	48209	19487.37	186	5651102.98	.046(.046-.047)	.962	.037	58493.61
Grade 11	46499	20282.65	188	5471124.61	.048(.047-.048)	.959	.035	62196.72
Grade 12	41095	18647.02	188	4832647.49	.049(.048-.049)	.957	.033	62409.30

3.3 Discussion

We found that the VPR model provided the best fit to the Project TALENT battery of the three models that were tested. Three studies now support the conclusion that the VPR model provides a better description of the structure of intelligence than the CHC model or its precursor the three-stratum model (Johnson & Bouchard, 2005b, Johnson, te Nijenhuis, & Bouchard, 2007), and four studies now support the VPR model over the Gf-Gc model (Johnson & Bouchard, 2005a, 2005b;

Johnson, Te Nijenhuis, et al., 2007). No study to date has contradicted this conclusion.

Nonetheless, the PT dataset had some limitations for testing the structure of intelligence differences. As always, test selection was of relevance to how the different theoretical models were represented. The PT test battery was not suited to testing all of the differences between the models because it lacked tests in certain domains such as long-term storage and retrieval (Glr) and reaction and decision speed (Gt) (McGrew, 2009). However, in its favor, the PT battery was not constructed according to any particular theory of intelligence, and, as we argue below, it was suitable for testing each model's predictions in mechanical/scientific and verbal/educational domains. A second potential limitation of our findings is that even though the sample was representative of American high school students in the early 1960s, the pattern of results may not generalize to more recent samples due to cultural and educational changes. The structure of ability may also shift with age, becoming more differentiated as students have the opportunity to learn more specialized knowledge and subsequently enter the workforce as adults (Li et al., 2004).In this section, we first discuss some possible reasons why the VPR theory outperformed CHC/Gf-Gc theory in explaining the structure of the PT battery. Second, we describe some differences in how the VPR model was manifested in PT compared to previous studies. Third, we turn to the theoretical implications of this study and previous comparison studies of the structure of intelligence.

3.3.1 The three theories in Project TALENT

The VPR model described the overall structure of the PT battery more accurately than the CHC and Gf-Gc models. Yet one possible disadvantage of the battery for the CHC model may have been that it effectively lacked second-order factors (such as Gc, Grw, etc.; see Figure 3.3) because there were not enough indicators for them. Nevertheless, CHC theory dictates that the six first-order factors we observed all load onto different second-order factors; thus the creation of second-order factor that were combinations of different first-order factors would not have been consistent with the theory (McGrew, 2009). In addition, such a model with only one higher-order factor received support from Carroll (1993), who found this structure in the male-only PT datasets SHAY01 and FLAN01. Such a model has been used as representative for CHC theory in PT by later researchers (Reeve, 2004; Reeve et al., 2006) Although Carroll (1993) found three second-order factors in the female-only dataset SHAY02 (see Tables 15.4 and 15.14, Carroll, 1993), two of these factors were inconsistent with both his three-stratum theory and the more recent CHC theory. These factors were the 2H factor (the combination of fluid and crystallized intelligence), which is lacking in both models, as is the Technical Information factor that received loadings from the General Information factor (K0), Math knowledge (KM) and General Science Information (K1). Thus, our version of the CHC model was consistent with the structure Carroll found in the male PT datasets, and was true to the most recent description of the CHC model in terms of the classification of first-order factors as loading onto particular second-order factors (McGrew, 2009). Nevertheless, our results contradicted the prediction derived from CHC theory that the six first-order factors in PT belong to six different domains (and thus make six independent contributions to *g*; see Figure 3.3). Similarly, our results also contradicted the four second-order factors of the Extended Gf-Gc model: crystallized intelligence (Gc), visualization (Gv), quantitative abilities (Gq) and cognitive speed (Gs); see Figure 3.2. Some of the second-order factors of the Gf-Gc and CHC models were likely not fully representative of the broad abilities in the models due to the relative narrowness of the PT test battery in certain domains (such Gs or Gq), but if the structure of ability were divided into those domains, rather than those predicted by the VPR model, then the models should still have demonstrated better fit than the VPR model. These same arguments can also be extended to the three models in the narrow selection of tests.

As noted above, we believe that, despite limitations for testing these theories in terms of the test battery content, PT was suitable for testing their predictions about tests in the mechanical/scientific and verbal/educational domains. An important area where the VPR model and CHC/Gf-Gc models differ is in their conceptualizations of where tests of mechanical/science knowledge fit in the structure of intelligence. The VPR model predicts that performance on these tests depends on the underlying Perceptual and/or Image Rotation abilities and skills of the test taker, as well as interest in and experience with the subject area. Thus the VPR model specifies that tests of mechanics and science information are distinct from measures of acculturated knowledge that rely more directly on verbally expressed knowledge (e.g. General Information). The Gf-Gc model does not make this distinction and instead holds that all measures of acquired knowledge are part of crystallized intelligence (Gc) as distinct from non-learned reasoning and information processing capacities. Although the CHC model does distinguish factors of domain-specific knowledge from Gc, it does not place specific mechanical/science knowledge with spatial ability as the VPR model does. The PT battery provided a difficult test for the VPR model because the mechanical and scientific information tests were based entirely on the recall of stored knowledge, and the Spatial/Reasoning factor was made up of tests that were nearly all visual-spatial and based on novel problem-solving. The results of the current study thus provide support for the VPR view that spatial ability and mechanical/science knowledge are linked under Perceptual ability.

Another area where the models differ is on how tests of English and Math ability fit into the intelligence hierarchy. Both the Gf-Gc and CHC models dictate that Quantitative abilities and English language abilities load onto separate second-order factors (Horn & Blankson, 2005; McGrew, 2009). The VPR model, in contrast, proposes that Math and English abilities are linked under the broad Verbal factor (though Math abilities also often load on the Perceptual factor, see Johnson & Bouchard, 2005b). Thus, the VPR model can also explain why a first-order English/Math factor was obtained in the PT battery, but the Gf-Gc and CHC model cannot.

Despite favoring the VPR model, the difference in fit according to the CFI and SRMR were not large, particularly for the narrow selection. The interpretation of the VPR as better fitting was dependent on the validity of a handful of cross-loadings. It is possible that proponents of the Gf-Gc and CHC theories could propose alternative better-fitting specifications of their models in Project TALENT. The VPR model also benefitted from a more flexible theoretical stance (see section 3.3.3 for further discussion).

In summary, the better fit of the VPR model may be explained because it made allowance for the underlying relation of the Spatial ability and

Mechanical/Science factors in the form of the second-order Perceptual factor, and the relation between the English and Math factors in form of the second-order Verbal factor. The lack of these factors in the CHC/Gf-Gc models and their too precise specification of second-order factors likely explain their poorer fits to the data.

3.3.2 Variations in VPR model specifications

The specifications of the VPR model in the studies that have fit it have varied. One notable difference in the PT VPR models from previous ones was that model fit was improved when the Information factor was separated from the broad second-order Verbal factor. Previous studies have only found only one Verbal factor in the VPR model (Johnson & Bouchard, 2005a, 2005b; Johnson, Te Nijenhuis, et al., 2007). Separation of the two factors may have been helpful because there was a surplus of Information and English/Math-related tests in this school-oriented battery, creating a spuriously high correlation between g and the Verbal factor with the Information loading upon it. This suggests that these two factors were bloated specific (overrepresentation of tests in closely related subject areas), as indicated by the high number of tests loading upon them relative to the other factors (see Tables 3.5 and 3.8). At the same time, the finding of more than one second-order Verbal factor is not inconsistent with VPR theory, as, like Vernon (1961), it proposes that the number of factors at any stratum and the precise borders between them are functions of the specificity of the tests in the battery rather than inherent facts of nature (see section 3.3. below for further discussion).

In the first study on the VPR model, the first-order Perceptual Speed factor was subsumed under the second-order Perceptual factor (Johnson & Bouchard, 2005b). Here, however, the first-order Speed factor was found to correlate highest with the English/Math factor, and thus was placed under the Verbal factor. These findings support Vernon’s (1961) view that speed factors can be subsumed by either of the Verbal or Perceptual factors, depending on the overall battery test content. In the PT test battery, the Speed factor was largely made up of tests with verbal/numerical content (e.g. Table Reading, Clerical Checking, Disguised Words). This factor can thus be compared more closely to the Fluency factor found in the operationalization of the VPR model in the Thurstone and Thurstone (1941) test battery (Johnson and Bouchard, 2005b). In the VPR model specified in Wolff and Buiten's (1963)'s battery, in contrast, the Speed factor was more perceptual in nature, involving more visual than verbal/numerical stimuli (Johnson, te Nijenhuis, & Bouchard, 2007). Despite its loading on the Verbal factor, the Speed factor in the current study did potentially also have a Perceptual component, as illustrated by the cross-loading of the Object Inspection test on the Spatial/Reasoning factor (see Tables 3.2 and 3.3). Thus, different factors that are each labelled as speed across studies may have quite different characteristics depending on the content of the tests involved in the factor.

The factors underlying the Perceptual factor in the PT battery can be compared most closely with those obtained in the analysis of de Wolff and Buiten (1963)'s test battery (Johnson, te Nijenhuis & Bouchard, 2007). In that battery, a first-order factor of mechanical reasoning was formed from tests of knowledge about tools and reasoning using mechanical principles, which is comparable to the Mechanical/Science factor we obtained in the broad selection of PT tests. Similar factors of mechanical reasoning were not identified in the two other studies on the VPR model because these test batteries lacked tests in that domain, with the exception of the mechanical ability test in the Minnesota Study of Twins Reared Apart (MISTRA) battery, which loaded onto the spatial factor (Johnson & Bouchard, 2005b). A second similarity with the Perceptual factor in de Wolff and Buiten (1963)'s battery is that its highest loading was from a factor of inductive reasoning formed mainly by matrix reasoning tests, which is similar to the Spatial/Reasoning factor of the current study. One remaining difference between the VPR model in the PT and previous studies is the lack of a separate Image Rotation factor, but this can be explained by the deficiency of pure tests for this ability in the PT battery aside from the Visualization in Three Dimensions test.

3.3.3 Theoretical implications for the structure of intelligence

An important aspect of VPR theory that distinguishes it from other theories of intelligence is the proposal that the main dimension, after *g*, along which the structure of intelligence is organized is the Verbal-Perceptual-Image Rotation dimension. Johnson and Bouchard (2005a) postulated that this dimension reflects the fact that Verbal abilities rely on the serial, analytic functions of the left brain hemisphere, while Perceptual and Image Rotation abilities rely more on the non-verbal, parallel functioning of the right hemisphere (see also Gustafsson, 1984). In further support of this view, Johnson and Bouchard (2007) found that when g was partialed from from the 42-test MISTRA battery, there was a residual dimension with Verbal abilities on one pole and Image Rotation and Perceptual abilities on the other. This dimension also displayed a strong sex difference. Further research has suggested that position on the Verbal-Image Rotation dimension is associated with regional brain differences (Johnson, Jung, Colom, & Haier, 2008).

McGrew (2009) observed that while the CHC framework has already had a strong influence on applied fields, "the adoption of the CHC umbrella term has been much slower in theoretical fields, such as research published in the journal *Intelligence*." (p. 3). One possible reason that basic researchers have been slow to embrace it is that, in contrast with VPR theory, CHC theory lacks the theoretical content and predictions to satisfy researchers' requirements for a basic theory of intelligence. Carroll (1993) stated that a theory about the structure of cognitive abilities should "provide hypotheses about the sources of individual differences in these abilities" (p. 631); however, he proposed or tested few such hypotheses for his three-stratum theory, and subsequent CHC researchers, as well as Extended Gf-Gc researchers, have seemed satisfied to produce ever-growing catalogues of ability factors, without substantial theoretical investigation into the cognitive, genetic or neurological underpinnings of these new factors (McGrew, 2009; Horn & Blankson, 2005).

One strong theoretical prediction of VPR theory, derived from its dimension view of intelligence, is that any attempt to specify the number of factors at any order or stratum is inappropriate because the dimension is continuous. The implication of this view, which is different from the discrete-factors view of the CHC and Gf-Gc theories, is that there are many points along the Verbal-Image Rotation dimension where factors could be found if appropriate tests were used. Thus, the only prediction for the content of factors is that they lie somewhere along this dimension, with their specific character dependent on the content of the tests in the battery.

Nonetheless, VPR theory would not predict, for example, that a factor would be found which combines Verbal and Image Rotation tests, because these represent positions that are far apart on the dimension. Another related prediction of VPR theory is that even the number of strata is dependent on the level of detail of measurement, and thus the number of strata could potentially be increased indefinitely with tests of sufficient breadth and detail. These theoretical points of view contrast with those in the Gf-Gc and CHC theories that the structure of intelligence is organized into a limited number of identifiable factors and strata.

Another important theoretical difference is that in the CHC and Gf-Gc models the broad second-order factors are conceptualized to depend more on either purportedly novel processing abilities or acquired knowledge (Carroll, 1993; McGrew, 2009; Horn & Blankson, 2005), whereas the VPR model maintains that Verbal, Perceptual and Image Rotations abilities all involve stored knowledge, which includes knowledge about how to approach problems involving freshly-generated processing, with the balance depending on the particular constructions of the tests and the experiences and knowledge of the test takers. For example, the Perceptual factor in the PT VPR model is incongruent with CHC and Gf-Gc theory because it received loadings from a factor made up from information tests (Mechanical/Science), and a factor made up of (often-presumed novel) reasoning tests (Spatial/Reasoning), but VPR theory holds that these factors are linked because of their reliance on similar spatial/mechanical content and underlying perceptual ability. This concept seems to have been overlooked in the literature, and in test design, because the distinction between crystallized and fluid factors is often confounded with the division between tests involving verbal and spatial content. Carroll (1993)'s classifications are also subject to this criticism. Of the factors classified as Gf in his review, the three most frequent first-order factors that had high loadings on Gf were Induction, Visualization and Sequential Reasoning (Carroll, 1993, p. 598). Induction was defined most often by Ravens Progressive Matrices, Visualization was made up entirely of loadings from visual-spatial tests, and Sequential Reasoning was defined most often by a test called Ship Destination, which involved simple calculations based on a diagram of ship locations (Table 6.1, Carroll, 1993). The first-order factors that Carroll (1993) most frequently proposed to load onto Gc were Verbal Ability, Language Development, and Reading Comprehension, all factors which were dominated by verbal subject-matter content (p. 599). A more recent example of the conflation of spatial and verbal factors with fluid and crystallized factors is given by Benson, Hulac and Kranzler (2010). Benson et al. (2010) found that the CHC model provided a better fit to the Fourth Edition of the Weschler Adult Intelligence Scale (WAIS-IV) than the structure proposed by the test designers. However, in the CHC model the Gc factor was defined solely by loadings from tests of verbal subject-matter, and the Gf factor received loadings from only visual-spatial tests plus the Arithmetic test.

In this study, we found again that the VPR model, which contains factors of verbal content and perceptual-spatial content that transcend the fluid/crystallized distinction, yielded better fit than models specifying a separate crystallized factors (such as Gc and Gkn) and fluid factors (such as Gv and Gs). This finding thus provides further evidence that the higher-order structure of intelligence is organized along a Verbal-Perceptual-Image Rotation dimension rather than characterized by broad factors distinguished mainly by their purported reliance on novel processing or acquired knowledge.

3.4 Linking cognitive ability with personality

Establishing the VPR model as the best-fitting model among the three most supported in the literature enabled exploration of the associations among cognitive ability and personality and occupational interests, using the VPR model as the basis for consideration of cognitive ability. The first step in this process was to examine personality-intelligence associations.

The PT data provided an opportunity to look not just at linear associations but also nonlinear associations, which require a large sample due to their relatively small effect sizes. Some research in the gifted literature provided a basis for hypothesizing that such nonlinear associations could be substantive at the right tail of the intelligence distribution. Hypotheses for the linear associations were drawn primarily from reviews of previous personality-intelligence research.

The analysis did not include specific abilities because there have been fewer studies of specific abilities and personality, hence making hypotheses in this area more difficult to formulate. A second reason was to keep the number of analyses manageable. As in the first study, all eight PT grade and gender samples were used.

Chapter 4: Linear and nonlinear associations between general intelligence and personality

4.1 Introduction

Intelligence and personality are important predictors of behavior and outcomes in many domains, notably in educational and occupational settings (Barrick & Mount, 1991; Hunt, 2011). In addition, there are some associations between intelligence and personality traits (Ackerman & Heggestad, 1997; DeYoung, 2011). Within the Big Five framework, general intelligence (*g*) is most strongly associated with Openness to Experience (r = .33 in the N-weighted meta-analysis of Ackerman & Heggestad, 1997). This connection may seem obvious since measures of Openness to Experience typically include items assessing engagement in intellectual pursuits, and because intelligence has often been held to be the cognitive part of personality (Cattell, 1950; DeYoung, 2011; Guilford, 1959). Nonetheless, intelligence is also related to personality traits that are considered the least cognitive, such as Extraversion and Neuroticism (DeYoung, 2011). Neuroticism has consistently shown modest negative correlations with general intelligence (r = -.15 in Ackerman & Heggestad, 1997), and most recent studies (performed after the year 2000) have shown that Extraversion also has a small but significant negative association with *g*, in the range of r = -.04 to -.11 (Luciano, Leisser, Wright, & Martin, 2004; Moutafi, Furnham, & Paltiel, 2005; Soubelet & Salthouse, 2011; Wolf & Ackerman, 2005). In addition, DeYoung (2011) found that in 9 studies not included in Ackerman and Heggestad's (1997) meta-analysis, Conscientiousness had a mean N-weighted correlation of -.12 with intelligence.

Nevertheless, some researchers have argued that the theoretical implications of these personality-intelligence correlations are limited due to their small size or inconsistency across studies (Eysenck, 1994; Soubelet & Salthouse, 2011; Zeidner, 1995). One possibility is that some intelligence-personality associations could be nonlinear, and thus missed by traditional linear analyses (E. J. Austin, Deary, & Gibson, 1997; E. J. Austin et al., 2002; Eysenck & White, 1964; Reeve et al., 2006). Findings in this area have, however, have often been negative. Austin et al. (1997)

found evidence for positive quadratic effects of Neuroticism and Openness to Experience on intelligence in one sample, but Austin et al. (2002) did not find any significant effects of this kind for the Big Five and Eysenck's Big Three in four other datasets. There are three theoretical and methodological issues surrounding these results.

First, different theories make alternative causal predictions about personality-intelligence relations. For example, Ackerman's PPIK theory (intelligence-as-process, personality, interests, and intelligence-as-knowledge) predicts that intelligence becomes related to personality through cognitive investment in four trait complexes which involve different personality traits and interests (Ackerman, 1996; Ackerman & Beier, 2003b). Alternatively, Chamorro-Premuzic and Furnham (2006) proposed that personality-intelligence relations can be conceptualized as the influence of personality traits on intellectual competence, where intellectual competence is defined as "an individual's capacity to acquire and consolidate knowledge throughout the life span" (p. 259, Chamorro-Premuzic & Furnham, 2006). PPIK theory suggests that cognitive factors causally contribute to broader constellations involving personality and interests (trait complexes), and thus that the association between all the variables is an emergent property due to reciprocal causation between all three variables. In contrast, Chamorro-Premuzic and Furnham's theory proposes that personality traits directly influence the development of intelligence. A third possibility is that intelligence contributes directly to the development of personality through conscious perceptions of adaptive benefit of particular behaviours, or through the influence of intelligence on motivations.

When estimating only linear effects, it is difficult to distinguish these possibilities without a longitudinal design, because effects are typically symmetrical no matter which ways the causal arrows are drawn. However, nonlinear analyses can pick up larger effects in one direction (e.g. there might be a quadratic effect of intelligence on Extraversion but no quadratic effect of Extraversion on intelligence), which can suggest that causal forces operated in this direction. Previous studies of quadratic effects have focused on the quadratic effects of personality on intelligence

(E. J. Austin et al., 1997; E. J. Austin et al., 2002); however, in the current study we assessed quadratic effects of intelligence on personality.

The second issue surrounding nonlinear personality-intelligence relations is that previous nonlinear studies were not performed with latent variables as predictors but with observed scores (E. J. Austin et al., 1997; E. J. Austin et al., 2002; Reeve et al., 2006). This limited their power because the size of the quadratic effect was not corrected for unreliability. Quadratic terms are particularly sensitive to unreliability of the predictor variable (Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009). Methodological researchers have observed that even using factor scores for the predictor can produce biased estimates of structural model parameters due to residual measurement error (D. Bartholomew, 1987; Harring, Weiss, & Hsu, 2012). Recently, Harring et al. (2012) found that, compared with methods that model latent quadratic terms directly, the use of factor scores led to substantial underestimation of quadratic coefficients.

A third issue is that to detect quadratic effects with small effect sizes, large sample sizes are needed. Under simulation, Harring et al. (2012) showed that for a medium-sized quadratic effect that accounted for 5% of the variance, even a small sample size of 50 was sufficient to obtain power over .80. However, in practice, quadratic or interaction effects can be considerably smaller than this, accounting for only 1% or 2% of the variance. To find these effects, very large sample sizes (i.e. of 500 or greater) are necessary. For example, Moosbrugger et al. (2009) found that for a quadratic effect size of 2% and a sample size of 400, average power was only 76% using latent estimation (power would be less with non-latent methods). Thus, the sample of Austin, Deary & Gibson (1997) and two of the four samples in Austin et al. (2002) may not have had sufficient power to detect small quadratic effects.

Reeve, Meyer and Bonaccio (2006) conducted one study on personality-intelligence relations that was sufficiently powered. Their study is directly relevant to ours as we made use of the same sample so we review their analysis in detail. Reeve et al. (2006) used a subsample of data from Project TALENT (PT), a nationally-representative study of approximately 400,000 American high school students in 1960. The sample in Reeve et al. (2006) consisted of 71,887 students in their final year of high school (seniors), with a mean of age 17.2 years (SD = 1.3). The ten PT personality scales were developed specifically for the Project in the late 1950s, before there was much consensus about models of personality structure. The scales used thus do not correspond directly to the Big Five framework in common usage today, but Reeve et al. (2006) related the scales to the Big Five by two methods. First, the three authors independently examined each scale's content and compared it to the content of the NEO-PI-R scales (P.T. Costa & MacCrae, 1992), and second, they re-administered the PT personality scales and IPIP scales for the Big Five to a sample of 219 college students. Table 4.1 summarizes the NEO-PI-R facet with which each PT scale was most closely associated (by rater consensus), as well as with which Big Five trait(s) the scales loaded in a joint factor analysis with IPIP scales (Reeve et al., 2006).

These relations provided a way to link the PT scales to the larger literature on personality-intelligence relations, which has frequently been organized according to the Big Five (e.g., Ackerman & Heggestad, 1997; Austin et al., 2002). The facet-matching by Reeve and colleagues may be limited due to imperfect content overlap, but the majority of the PT scales displayed good convergent validity with the Big Five factors predicted to subsume them (factor loadings = .42 to .81). In addition, the content of the PT scales was facet-like; hence they could be viewed as analogous to facets of the Big Five, with the exception that some scales (e.g. Self-Confidence) would be facets of more than one Big Five factor.

Table 4.1
Associations of the Project TALENT personality scales with the Big Five.

PT scale	NEO-PI-R facet	Big Five trait loading(s)[a]
Sociability	Gregariousness (E)	E (0.69), A (0.38)
Calmness	Anger (ES) - reversed	ES (0.69)
Vigor	Activity (E)	E (0.43)
Social sensitivity	Sympathy (A)	A (0.81)
Tidiness	Orderliness (C)	C (0.79)
Culture	Aesthetics (O)	O (0.51) A (0.44)
Self-confidence	Self-consciousness (ES) - reversed	E (0.60) ES (0.60)
Mature personality	Achievement Striving (C)	C (0.63) A (0.35)
Impulsiveness	Cautiousness (C) - reversed	E (0.42)
Leadership	Assertiveness (E)	E (0.51) O (0.41)

[a] Loadings obtained in a joint factor analysis of the IPIP and PT scales by Reeve et al. (2006)

Reeve et al. (2006) found that *g* correlated positively and substantively (above .15 in their definition) with the scales Mature Personality, Calmness, and Self-Confidence in grade 12 males. These correlations were also observed in grade 12 females, where Culture and Social Sensitivity were also correlated positively with *g*. These associations may, however, have been influenced by measurement artefacts because the PT personality scales were nearly uniformly positively correlated with each other. The mean of the inter-scale correlations in the senior sample was .38 in males, and .35 in females (SD = .14 in both samples). Reeve et al. (2006) did not address this common variance among personality scales (similar factors in other personality inventories have been termed 'general factors of personality'; Rushton & Irwing (2008). This common variance was relevant because it correlated positively with *g* in Project TALENT (mean r = .28 in all samples), and thus we predicted that it would affect the correlations of the scales with *g*.

Recent research has suggested that the common variance between Big Five measures is in large part due to rater bias. In a meta-analysis of 45 multi-trait multi-method samples, Chang, Connelly and Geeza (2012) found that much of the common variance between Big Five personality scales is due to method variance specific to raters, which likely includes response biases such as socially desirable responding. After rater effects were controlled for in the CTOM (correlated traits, orthogonal methods) model, adding a general factor of personality (GFP) above the Big Five factors resulted in a substantial decrement in model fit compared the model allowing free covariance between the Big Five (Chang et al. 2012). Moreover, the GFP had non-substantive loadings from Extraversion (.03) and Openness to Experience (-.09), supporting the view that there is no single factor that sits above the Big Five in multi-informant data (however, a model with Digman's Alpha and Beta were still found to be plausible by Chang et al., 2012). A number of studies have now supported the conclusion that the GFP emerges for artifactual methodological reasons (Anusic, Schimmack, Pinkus, & Lockwood, 2009; M.C. Ashton, Lee, Goldberg, & de Vries, 2009; Bäckström, Björklund, & Larsson, 2009; de Vries, 2011).

As detailed further below, we also observed that several of the PT scales in Reeve et al. (2006), and in our initial analysis, displayed stronger positive correlations with *g* than were expected based on such correlations in the Big Five, with which the PT scales correlated. This, combined with the moderate correlation of the GFP to *g*, suggested that common variance may have acted as a confounder in the estimates. Because we were primarily interested in the relations of the individual scales to *g*, and wished to err on the side of under-estimation rather than over-estimation, we conducted our analyses while controlling for the GFP.

In addition to linear associations, Reeve et al. (2006) looked for nonlinear relations by converting the personality scores into extremeness scores ($X - Mean_x|$) and examining their correlations with *g* factor scores. Reeve et al. (2006) did not observe any correlations between the extremeness scores and *g* above a selected cut-off of .15. However, there were two limitations to their method of looking for quadratic effects. First, whereas extremeness scores may suggest the presence of quadratic trends, they are not equivalent to examining true quadratic effects which predict scores with the form $|X^2\text{-}Mean_x|$. Second, Reeve et al. (2006) chose to convert the personality scale scores rather than the intelligence test scores in PT to extremeness scores, thus examining the effect of extreme personality on intelligence (Rosenthal & Rosnow, 1991). This is the same direction of effect investigated by Austin and colleagues (Austin et al., 1997; Austin et al., 2002). As noted, we were instead interested in examining the effects of intelligence on personality. This had the added advantage of greater power, due to the greater reliability of the latent *g* factor compared to the observed personality scales.

The aim of our study was to re-examine linear and nonlinear relations between *g* and personality in Project TALENT, taking into account common variance among the scales. Moreover, we used structural equation modeling (SEM), which avoids using factor scores and allows for direct estimation of latent linear and quadratic effects (Klein & Moosbrugger, 2000). In addition to SEM, we used generalized additive models (GAMs; Hastie & Tibshirani, 1986) to explore further possible nonlinear trends. The PT data were suited to our aims because of its large and relatively population-representative sample of nearly 400,000 high school students in four grades, allowing for the possibility of replication across grade subsamples.

4.1.2 Linear personality-intelligence associations in Project TALENT

We focused on the personality-intelligence literature (primarily on the Big Five) in generating our hypotheses about specific associations.

Openness to Experience displays a positive correlation with *g* (Ackerman, 2009; Ackerman & Heggestad, 1997; DeYoung, 2011), and the two PT scales that loaded significantly on Openness to Experience were Culture and Leadership (Reeve et al., 2006). Neither scale is a pure measure of Openness/Intellect (see Table 4.1); therefore, we hypothesized that their correlations with *g* would be positive, but smaller in size than the .33 value in meta-analysis of Ackerman & Heggestad, 1997.

Five of the ten scales in PT had primary loadings on Extraversion, which has typically shown small negative associations with intelligence (Wolf & Ackerman, 2005; Moutafi, Furnham & Paltiel, 2005; Austin et al., 2002). However, this relation is not uniform for all facets of Extraversion. Ackerman and Wolf (2005) suggested that Extraversion should be split to reflect two different aspects: social closeness (the need for intimacy) and social potency (the need for making an impact on others). They also hypothesized that "Individuals high on social closeness may be less likely to invest their time in intellectually engaging tasks, leading to lower scores on intelligence tests" (p. 533, Wolf & Ackerman, 2005). Partially consistent with this, their meta-analysis of 48 samples showed that the correlation between social potency and intelligence was slightly positive ($r = .04$, $p < .05$), whereas the intelligence association with social closeness was not significantly different from zero ($r = -.01$) (Wolf & Ackerman, 2005). Similarly, Pincombe, Luciano, Martin & Wright (2007) found that the excitement-seeking and gregariousness facets of NEO Extraversion correlated negatively with IQ ($r = -.09$ and $r = -.15$, respectively). We thus anticipated that PT Sociability and Impulsiveness scales would show negative associations with intelligence (due to their face-value relations with social closeness and excitement-seeking), whereas Vigor, Self-Confidence and Leadership would show positive associations (due to their face-value relations with social potency).

The Big Five trait Neuroticism has a negative correlation with intelligence (Ackerman & Heggestad, 1997; DeYoung, 2011). Based on their face-value contents, and the findings of Reeve et al. (2006), the PT scales of Calmness and Self-

Confidence represent the converse of Neuroticism (Emotional Stability); therefore, we predicted these scales would display positive associations with intelligence.

The literature has suggested that Big Five Conscientiousness has a small negative association with intelligence (Moutafi, Furnham & Paltiel, 2005; DeYoung, 2011). In addition, in a sample of British adults, Moutafi, Furnham and Crump (2003) found that the Orderliness facet of Conscientiousness in particular had a negative correlation with *g* (r = -.18), which they argued may be because lower-intelligence individuals use planning and organization to compensate for their disadvantage on intellectual tasks (see also Chamorro-Premuzic and Furnham, 2006). The PT scale Tidiness was related on a content basis to Orderliness by Reeve et al. (2006); therefore, we hypothesized that it would have a negative correlation with intelligence. The PT scale Mature Personality was also related to Conscientiousness by Reeve et al. (2006), hence we predicted a negative association for it.

Big Five Agreeableness has typically not been found to have significant correlations with intelligence (Ackerman & Heggestad, 1997; DeYoung, 2011); hence we did not make any directional hypothesis regarding the PT scale Social Sensitivity, which was the only PT scale with a high correlation with Agreeableness according to Reeve et al. (2006).

4.1.3 Possible nonlinear associations

Although nonlinear associations between intelligence and personality have rarely been found, some suggestive evidence for nonlinear associations has been found in research on gifted children and adolescents. This has primarily been done with the Myers-Briggs Type Indicator (MBTI; Myers, McCaulley, & Most, 1985).

A meta-analysis of 14 studies of gifted adolescents, mostly identified through talent searches using the SAT and selection into gifted programs, showed that they were substantially more likely to fall on one side of the dichotomous MBTI dimensions than a norm group of students (Sak, 2004). Gifted adolescents were more likely to select Introversion over Extroversion (48.7% compared to 35.2% in the non-gifted sample), Intuition over Sensation (71.6% compared to 31.9%), and Perceiving over Judging (60.1% compared to 45.4%), as well as marginally more likely to prefer Thinking to Feeling (53.8% compared to 47.5%; Sak, 2004). Studies in adults have found that MBTI Extroversion to be strongly related to Big Five Extraversion (r = .74), whereas Intuition is strongly related to Openness to Experience (r = .72); MBTI Thinking and Perceiving are moderately negatively correlated with Agreeableness (r = -.44) and Conscientiousness (r = - .49), respectively (correlations for the male sample in R. McCrae & Costa, 1989). Therefore, by extension it can be predicted that gifted adolescents may be substantially higher in Openness to Experience and lower on Extraversion, Agreeableness and Conscientiousness than non-gifted adolescents. A recent study of Israeli adolescents, who were selected as the top 1% to 3% of performers on an intelligence test, confirmed this pattern for Openness to Experience (d = .51) and Agreeableness (d = -.28), and also showed that gifted adolescents were lower in Big Five Neuroticism than non-gifted adolescents (d = -.26) (Zeidner & Shani-Zinovich, 2011). Group differences in Conscientiousness and Extraversion were in the expected direction based on MBTI studies, but non-significant (Zeidner & Shani-Zinovich, 2011).

The presence of some substantial mean differences between gifted and non-gifted groups suggests that average personality level might differ to an expanding (e.g. exponential) degree with increasing ability level, although it is possible that linear effects could produce these effects as well. Exponential functions may most closely approximate differences in certain personality traits with increasing ability level, but such trends would also be captured by quadratic effects, at least for one side of the parabolic curve. One issue relating to this testing is that some studies have also found increases in personality variance with higher intelligence (e.g. in the MBTI; Myers & McCaulley, 1985). This may violate the assumption of homogeneity of variance underlying generalized linear models, although these models are robust to some level of heteroscedasticity (Tabachnick & Fidell, 2007). It is possible that higher cognitive ability is causally linked to increases in personality variance, as intelligence potentially facilitates more flexible adjustment of personality to the environment; however, in this study we focused on mean-level differences in personality.

Given the evidence in the gifted literature, we hypothesized that positive quadratic trends would be observed for the PT scales associated with Openness to Experience (Culture and Leadership) as well as Emotional Stability (Calmness and Self-Confidence). We also predicted that negative quadratic effects (an inverted-U shape) would be observed for the scales associated with the social closeness aspect of Extraversion (Sociability), Agreeableness (Social Sensitivity), and Conscientiousness (Tidiness and Mature Personality). Because less is known about personality in low-ability groups, these predictions were based on the trend for above-average intelligence.

4.2 Method

4.2.1 Sample

Project TALENT participants were obtained by a stratified random sample of all public and private high schools in the United States in 1960 (Flanagan et al., 1962). The PT dataset was thus a nationally-representative sample of approximately 5% of the student population. The full sample consisted of 376,213 students, with approximately 100,000 students in each grade from 9 through 12. Of the full sample, 50.13% was female. The age range was from a mean of 14.4 in grade 9 (SD = .78) to 17.3 in grade 12 (SD = .67). The full individual age range was 8 to 21.

4.2.2 Intelligence measures

The intelligence measures for the current study were selected from the PT aptitude and achievement tests, using the broad selection of 37 tests as defined in a previous study (for descriptions of the tests and reliabilities see Major, Johnson & Deary, 2012; see also Flanagan et al., 1962).

The data screening methods for the intelligence tests were the same as used by Major, Johnson and Deary (2012). Scores on the PT response credibility index, which was based on a screening test assessing illiteracy, mental disability or an apathetic testing attitude, were used to exclude participants who did not reach the cut-offs set by the PT study designers, except where only mental slowness was indicated (a low score for the number of responses on the Clerical Checking test). Transformations were applied to three tests that displayed non-normal distributions

(Capitalization, English Usage and Table Reading), and cases showing severe problems with multivariate outliers were removed (Major, Johnson & Deary, 2012). Following data screening of the intelligence tests, total sample size was reduced to 366,939 (2.47% of the sample removed, the vast majority for low screening scores).

4.2.3 Personality measures

The PT personality scale scores were derived from 108 items that asked students how typical certain personal attributes and behaviors were of them. Table 4.2 contains sample items for the scales; reliability coefficients from Reeve et al. (2006) are presented due to their lack of availability from the original study. The responses to personality items were on five-point Likert scale. The scores available in the PT dataset were scale scores, which were obtained by assigning a score of 1 to items where the student indicated that the item described them "extremely well" or "quite well" (the two most affirmative responses), and a score of 0 to other responses ("fairly well", "slightly", or "not very well"). The converse scoring method was used for negatively-phrased items (Wise et al., 1979).

Table 4.2
Personality test descriptives.

Scale	Sample item	Items	Reliability[a]
Sociability	"I like to be with people most of the time"	12	.83
Calmness	"I am usually self-controlled"	9	.81
Vigor	"I am full of pep and energy"	7	.76
Social sensitivity	"I never hurt another's feelings if I can avoid it"	9	.79
Tidiness	"I like to do things systematically"	11	.85
Culture	"I think culture is more important than wealth"	10	.69
Self-confidence	"I am usually at ease"	12	.79
Mature personality	"I make good use of all my time"	24	.90
Impulsiveness	"I usually act on the first plan that comes to mind"	9	.69
Leadership	"People naturally follow my lead"	5	.65

[a] Reliabilities from the sample of 219 college students in Reeve et al. (2006).

4.2.4 The general factor of personality

The raw PT personality scales displayed mean inter-correlations that ranged from .35 in the grade 12 females (SD = .14), up to a maximum of .42 in the grade 9 males (SD = .13). Across the eight samples (four grades by two genders), the first common factor accounted for a mean of 41.3% of variance (SD = 2.2%). Potential sources for this common variance included artifactual sources such as method variance (e.g. due to pencil-and-paper testing), acquiescence bias and socially-desirable responding, and non-artifactual true score variance.

Although it was not possible to disentangle these sources directly, some evidence suggested that this common variance was potentially confounding the relations of personality scales with *g*. Several personality scales displayed unexpectedly positive correlations with *g*. The Tidiness scale, which Reeve et al. (2006) identified on a content basis with the Orderliness facet of Conscientiousness, displayed a positive correlation with *g* in all samples (mean r = .16 in males, .10 in females; see Appendix B). This observation contradicted the finding of Moutafi et al. (2003) that Orderliness is negatively associated with *g*, and that Conscientiousness in general in also negative related (DeYoung, 2011). Similar inferences could be drawn for the Sociability and Impulsiveness scales, which were predicted to have negative associations with *g* based on the literature, but instead showed small positive correlations (Sociability: r = .09/.05 in males/females; Impulsiveness: r = .03/.10 for males/females)[8]. We hypothesized that the positive correlation between the GFP and *g* could account for these positive correlations (mean r = .28 in both males and females).

In order to aid in the interpretation of the GFP, we performed a re-analysis of the college sample data from Reeve et al. (2006)[9]. The general factor from the PT scales, extracted through maximum likelihood estimation, explained 25.9% of the variance in the college sample. The GFP was then correlated with the individual items (item-level data were not available in PT). The Vigor scale was over-represented in items that correlated most highly with the GFP: six of the seven items assessing Vigor were in the top 10 most highly-correlated items, including the most highly correlated item ("I am energetic", r = .63). The Vigor scale also had the highest loading on the GFP (.71) in the college sample. In addition to this trend, only 23 of the 108 personality items (21%) contained statements that referred to other people's views (e.g. "people consider me sociable"), but 8 of these items were in the

[8] The correlation between Impulsiveness and *g* in grade 9 males was non-significant, however.
[9] Data obtained through personal communication (September 18, 2012).

top 20 most correlated with the GFP (40%). This finding suggested that items that primed reputational concerns were more closely tied to the GFP. In the college sample, the GFP was most highly associated with items that seemed to tap the form of socially-desirable responding that has been termed egoistic self-enhancement (Paulhus & John, 1998). Nonetheless, this interpretation may not generalize entirely to the PT sample, as indicated by differences in the GFP loadings in the two samples. In the college sample a lower loading was seen particularly on Tidiness (.29, compared to .70 in PT). This finding indicated that Tidiness was more integral to the GFP in PT, suggesting that the GFP in PT had more to do with Conscientiousness than in the college sample.

Regardless of whether the correlation between the personality factor and *g* was artifactual or not, our primary interest was in the relations of the individual scales with *g*. Therefore, we decided to remove influence of the common variance from the scales. The scales were regressed onto the GFP, and residuals retained for the further analyses. To verify that the residualization did not damage the convergent validity of the PT scales we examined their correlations with the predicted IPIP Big Five scales in the college sample. Compared to the mean correlation of the unresidualized scales ($r = .56$, $SD = .13$), the mean correlation decreased to $r = .36$ ($SD = .19$). This reduction was consistent with the high correlation of the GFP in the PT scales with the GFP in the IPIP scales ($r = .77$). When the variance in the PT GFP that was explained by the IPIP GFP was removed (through regression) prior to using it to residualize the PT scales, there was no reduction in the correlation between the residualized PT scales and the IPIP scales (mean $r = .56$, $SD = .12$). Removing the PT GFP appeared to reduce the convergent validity of the PT scales, but this seemed to be because the IPIP and PT scales shared method or rater variance, captured by their GFPs, which inflated the initial correlations. The results of the analyses with the original scales are presented in the supplemental materials (Appendix B), but the focus of all further presentation is on the residualized scales.

Following residualization for the GFP, the personality scales were screened for normality and outliers. Outliers were capped at four standard deviations above and below the mean (approximately the most extreme score expected in our samples). The scales Impulsiveness and Leadership displayed positive skewness in all samples, therefore a square-root transformation (with reflection) was applied to them. Following these transformations, all personality scales displayed adequate normality (all skewness and kurtosis z values below 0.5). In contrast to the raw scales, the mean correlation between the residualized scales was slightly negative, and ranged from -.096 (SD = .12) in grade 11 females to -.099 (SD = .09) in grade 9 males. Table 4.3 displays the correlations in grade 10 males and females.

Table 4.3

Correlations among personality scales after removal of the general personality factor (frade 10 males/females).

	Sociability	Social Sensitivity	Vigor	Calmness	Tidiness	Culture	Self-confidence	Mature Personality	Impul-siveness	Leadership
Sociability	–	.019	.182	-.160	-.160	-.183	.109	-.360	.059	-.035
Calmness	-.037	–	-.169	-.106	-.189	-.120	-.201	-.278	-.007	-.150
Vigor	.114	-.187	–	-.186	-.189	-.183	-.002	-.150	.104	.020
Social sensitivity	-.133	-.106	-.165	–	-.112	-.213	.056	-.164	-.130	-.164
Tidiness	-.162	-.203	-.157	-.155	–	-.052	-.191	-.004	-.179	-.200
Culture	-.127	-.068	-.191	-.202	-.041	–	-.182	-.211	-.047	-.133
Self-confidence	.049	-.194	-.069	.082	-.116	-.143	–	-.081	-.012	.030
Mature personality	-.339	-.261	-.130	-.168	-.080	-.231	-.032	–	-.085	-.040
Impulsiveness	.065	-.037	.035	-.095	-.145	-.048	-.072	-.103	–	.116
Leadership	-.051	-.082	-.043	-.145	-.186	-.092	-.049	-.079	.118	–

Male correlations are below the diagonal, females above.

4.2.5 Methods of analysis

We searched for linear and nonlinear associations between *g* and the personality scales in two ways. The first method was to estimate linear and quadratic effects using latent moderated structural equation modeling (LMS; Klein & Moosbrugger, 2000). LMS directly models the quadratic term as the interaction of a latent variable with itself (or the square of the variable), and corrects for the multivariate non-normality of the term, making it a better method than regression (Harring, Weiss & Hsu, 2012; Moosbrugger, Schermelleh-Engel, Kelava & Klein, 2009). LMS was performed in Mplus 5.21.

Secondly, we ran generalized additive models (GAMs; Hastie & Tibshirani, 1986), using the R package 'mgcv' (Wood, 2006). A GAM is a generalized linear model in which the linear predictor depends on unknown smooth functions of the predictor variables. The smooth functions are represented by regression splines with a particular basis function (for our analyses, the cubic basis was selected). The degree of smoothing of the spline is determined by the generalized cross validation score, which is a measure of how well the spline fits across datasets with each datum left out in turn (see Wood, 2006, for more details). We used GAMs to explore other possible nonlinear trends apart from quadratic trends between the personality scales and *g*.

Using the LMS and GAM approaches, we estimated the effects of *g* on the ten personality scales in each of eight samples divided by grade and sex. We selected the direction of effect of *g* on personality because we preferred this direction theoretically and because *g* was more reliably measured than the personality variables. In addition, it was not possible to estimate latent personality traits because of the lack of item-level data. Thus, *g* as a predictor allowed for LMS estimation of quadratic effects.

The measurement model used for *g* was the VPR model, which has been shown to fit well to these data (Major, Johnson & Deary, 2012). The variance explained by each effect in the LMS models was obtained by subtracting the residual variance of the personality scales from 1 (as the personality scales were standardized). GAMs were estimated with *g* factor scores obtained from the VPR model.

Missing data were handled with through direct maximum likelihood estimation, which requires the assumption the data were missing at random (MAR). This assumption was tenable in PT because it is unlikely that students purposely avoided particular aptitude tests or the personality scales. In addition, only 2.3 to 3.2% of the ability test scores and 1.0 to 3.5% of the personality test scores were missing in each sample.

4.3 Results

Tables 4.4 and 4.5 display the standardized linear and quadratic effects of *g* on the residualized personality scales for males and females, respectively. Figure 4.1 (males) and Figure 4.2 (females) illustrate the predicted mean-level differences in personality based upon the estimated linear and quadratic effects in the grade 10 samples.[10] Social Sensitivity in males and Calmness in females were omitted from the figures due to the lack of significant linear or quadratic effects.

In the male samples (Table 4.4, Figure 4.1), the largest linear effects of *g* were on Sociability (beta = -.042 to -.130), Calmness (beta = .076 to .104), and Self-Confidence (beta = .106 to .131). Substantial negative quadratic effects (R^2 of approximately 2% or greater) were observed for Sociability (beta = -.146 to -.159) and Vigor (beta = -.107 to -.116). Positive quadratic effects were observed for Mature Personality (beta = .099 to .119) and Leadership (beta = .106 to .124).

In the female samples (Table 4.5, Figure 4.2), the largest linear effects of *g* were on Sociability (beta = -.077 to -.195), Tidiness (beta = -.064 to -.163), and Mature Personality (beta = .075 to .140). Substantial negative quadratic effects were seen on Sociability (beta = -.140 to -.155) and Tidiness (beta = -.064 to -.163), and a positive quadratic effect was found on Mature Personality (beta = .075 - .140).

[10] The figures are at the end of the Results section. A separate page for figure titles and captions is provided due to lack of space.

Table 4.4
Standardized linear and quadratic effects of *g* on the personality scales (males).

Trait	Linear effect				Quadratic effect			
	Gr. 9	Gr. 10	Gr. 11	Gr.12	Gr. 9	Gr. 10	Gr. 11	Gr. 12
Sociability								
Beta	-.042	-.087	-.118	-.130	-.146	-.148	-.155	-.159
R^2	.002	.007	.014	.017	.032	.034	.037	.039
Calmness								
Beta	.076	.094	.101	.104	–	–	–	–
R^2	.006	.009	.010	.011	–	–	–	–
Vigor								
Beta	.053	.020	–	-.022	-.114	-.116	-.107	-.110
R^2	.003	.000	–	.000	.019	.020	.017	.019
Social Sensitivity								
Beta	-.029	–	.017	.020	–	–	–	-.016
R^2	.001	–	.000	.000	–	–	–	.000
Tidiness								
Beta	–	-.020	-.058	-.086	-.072	-.069	-.084	-.088
R^2	–	.000	.003	.007	.008	.008	.011	.013
Culture								
Beta	-.087	-.073	-.082	-.067	.024	.054	.077	.093
R^2	.008	.005	.007	.005	.000	.004	.008	.011
Self-Confidence								
Beta	.131	.107	.106	.118	–	–	–	.026
R^2	.017	.011	.011	.014	–	–	–	.001
Mature Personality								
Beta	.066	.061	.070	.056	.119	.115	.113	.099
R^2	.004	.004	.005	.003	.022	.021	.020	.016
Impulsiveness								
Beta	-.125	-.076	-.032	-.023	.031	.021	.028	.018
R^2	.016	.006	.001	.000	.001	.000	.001	.000
Leadership								
Beta	-.116	-.094	-.048	–	.124	.113	.107	.106
R^2	.013	.009	.002	–	.023	.022	.017	.016

Effects greater than .015 were significant at $p < .001$, with no adjustment for multiple testing. Non-significant effects are not shown.

Table 4.5
Standardized linear and quadratic effects of *g* on the personality scales (females).

Trait	Linear effect				Quadratic effect			
	Gr. 9	Gr. 10	Gr. 11	Gr.12	Gr. 9	Gr. 10	Gr. 11	Gr. 12
Sociability								
Beta	-.077	-.115	-.161	-.195	-.153	-.155	-.149	-.140
R^2	.006	.013	.026	.038	.036	.037	.036	.031
Calmness								
Beta	–	–	–	–	–	–	–	–
R^2	–	–	–	–	–	–	–	–
Vigor								
Beta	.041	–	–	-.032	-.074	-.077	-.065	-.050
R^2	.002	–	–	.001	.008	.009	.007	.004
Social Sensitivity								
Beta	–	–	.017	–	-.052	-.047	-.079	-.075
R^2	–	–	.000	–	.004	.008	.010	.009
Tidiness								
Beta	-.064	-.105	-.137	-.163	-.092	-.100	-.106	-.120
R^2	.003	.011	.019	.027	.015	.016	.018	.023
Culture								
Beta	.032	.066	.071	.078	.030	.042	.045	.080
R^2	.001	.004	.005	.006	.001	.003	.003	.005
Self-Confidence								
Beta	.047	.040	.034	.055	–	.022	.039	.043
R^2	.002	.002	.001	.003	–	.000	.003	.003
Mature Personality								
Beta	.075	.083	.133	.140	.162	.166	.149	.122
R^2	.006	.007	.018	.020	.041	.042	.035	.024
Impulsiveness								
Beta	-.036	.046	.055	.075	.069	.079	.046	.046
R^2	.001	.002	.003	.006	.008	.010	.003	.003
Leadership								
Beta	-.111	-.109	-.079	-.031	.095	.093	.096	.106
R^2	.012	.012	.006	.001	.014	.013	.014	.017

Effects greater than .015 were significant at $p < .001$, with no adjustment for multiple testing. Non-significant effects are not shown.

Of our nine hypotheses about the linear effects of *g* on the personality traits, five were supported in all male samples (positive: Calmness, Self-Confidence; negative: Sociability, Tidiness, Impulsiveness), and one more received support in some grades (the positive effect on Vigor)[11]. In females, four hypotheses were supported (positive: Culture, Self-Confidence; negative: Sociability, Tidiness) and two had mixed support (the positive effect on Vigor and negative effect on Impulsiveness). The most unexpected finding was a negative linear effect of *g* on Leadership. This effect, in combination with the positive quadratic effect of *g* on

[11] The negative linear effect of *g* on Tidiness in grade 9 males was only significant at $p < .05$, and hence would not survive correction for multiple testing. Due to the effects in the other three samples, however, we counted this effect as significant across all samples.

Leadership, resulted in the highest levels of Leadership being observed for those with low *g* (see Figures 4.1 and 4.2). This finding may be less trustworthy than the others, however, because the Leadership scale only contained five items, and displayed borderline reliability (alpha = .65) in Reeve et al. (2006). We reserve interpretation of the meaning of the effects for the Discussion.

Of our eight predicted quadratic effects, four were supported in male samples (positive: Culture, Leadership; negative: Sociability, Tidiness), and six were supported in female samples (positive: Culture, Self-Confidence, Leadership; negative: Sociability, Social Sensitivity, Tidiness). The most important deviation from our hypotheses was for the Mature Personality scale, which was predicted to have a negative quadratic association with *g*, but instead had a positive one.

4.3.1 LMS results compared to GAM results

Figure 4.3 shows a comparison of the fitted functions in the LMS and GAM models for the example of Sociability in grade 10 males. As can be seen, the predicted personality levels are similar in both models. In general, visual inspection of GAM-predicted values showed a close correspondence with LMS results, indicating that a combination of linear and quadratic effects gave a good approximation of the relations revealed by the GAMs (other graphs of the GAMs are available from the first author). In addition, the R^2 for the GAMs were consistent with the variance explained by the combination of the linear and quadratic effects of LMS (slightly more variance was accounted for in the LMS models due to the use of a latent *g* factor instead of factor scores). For the GAMs in males, the three personality traits where *g* predicted the most variance were Sociability (3.3%), Leadership (2.1%) and Tidiness (1.7%). In females it was Sociability (4.4%), Mature Personality (3.7%), and Tidiness (2.5%). The variance explained in the personality scales was higher for females than males in a number of cases, although this varied greatly across scales (see Tables 4.4 and 4.5). Table B3 in Appendix B compares the AICs for the GAM models to the null models in the grade 10 samples. Compared with the null models, the GAM models all displayed lower AICs, indicating that they added predictive power compared to a model with no predictors (and were better fitting).

4.3.2 Grade and sex differences

The comparison of grade and sex differences in the estimated personality-intelligence relations requires the assumption of measurement invariance between the samples for intelligence and personality. This assumption could not be tested for the personality scales due to lack of item-level data. In addition, measurement invariance testing revealed that although configural and weak invariance were tenable across grade and sex the VPR intelligence model, strong invariance (equality of the intercepts) was not supported in both cases, as indicated by an decrease in CFI > .010 (Cheung & Rensvold, 2002). Therefore, the differences in personality-intelligence relations across samples must be interpreted with caution, as they may be attributable to differences in the measurement of intelligence (or personality).

With this caveat in mind, there were some differences in the measured relations across grade. Notably, the linear relation of Sociability with *g* was more negative with increasing grade level (comparing grade 9 to grade 12 in males: Δ beta = -.88, log-likelihood ratio test: $\chi 2$ (1) = 263.58, $p < .001$; in females: Δ beta = -.118, $\chi 2$ (1) = 514.98, $p < .001$). Two other important trends were the reduction of a negative association of Leadership with *g* at higher grades (in males, Δ beta =.106, $\chi 2$ (1) = 387.06, $p < .001$; in females, Δ beta = .080 $\chi 2$ (1) = 233.40, $p < .001$) and an increase of the negative association of *g* with Tidiness (in males, Δ beta = -.075, $\chi 2$ (1) = 189.34, $p < .001$; in females, Δ beta = -.099 $\chi 2$ (1) = 364.73, $p < .001$).

4.3.3 Figures 4.1 to 4.3: titles and captions

Figure 4.1 Mean personality as predicted by general intelligence (grade 10 males).

Caption: Personality scales and *g* are in standard units. Light lines represent 2 standard errors (SEs) above and below the mean (approximate 95% confidence interval). SEs obtained from GAM models.

Figure 4.2 Mean personality as predicted by general intelligence (grade 10 females).

Caption: Personality scales and *g* are in standard units. Light lines represent 2 standard errors (SEs) above and below the mean (approximate 95% confidence interval). SEs obtained from GAM models.

Figure 4.3 LMS and GAM-predicted sociability as a function of general intelligence (grade 10 males).

Caption: LMS estimate = solid grey line. GAM estimate = dashed line. Sociability and *g* are in standard units.

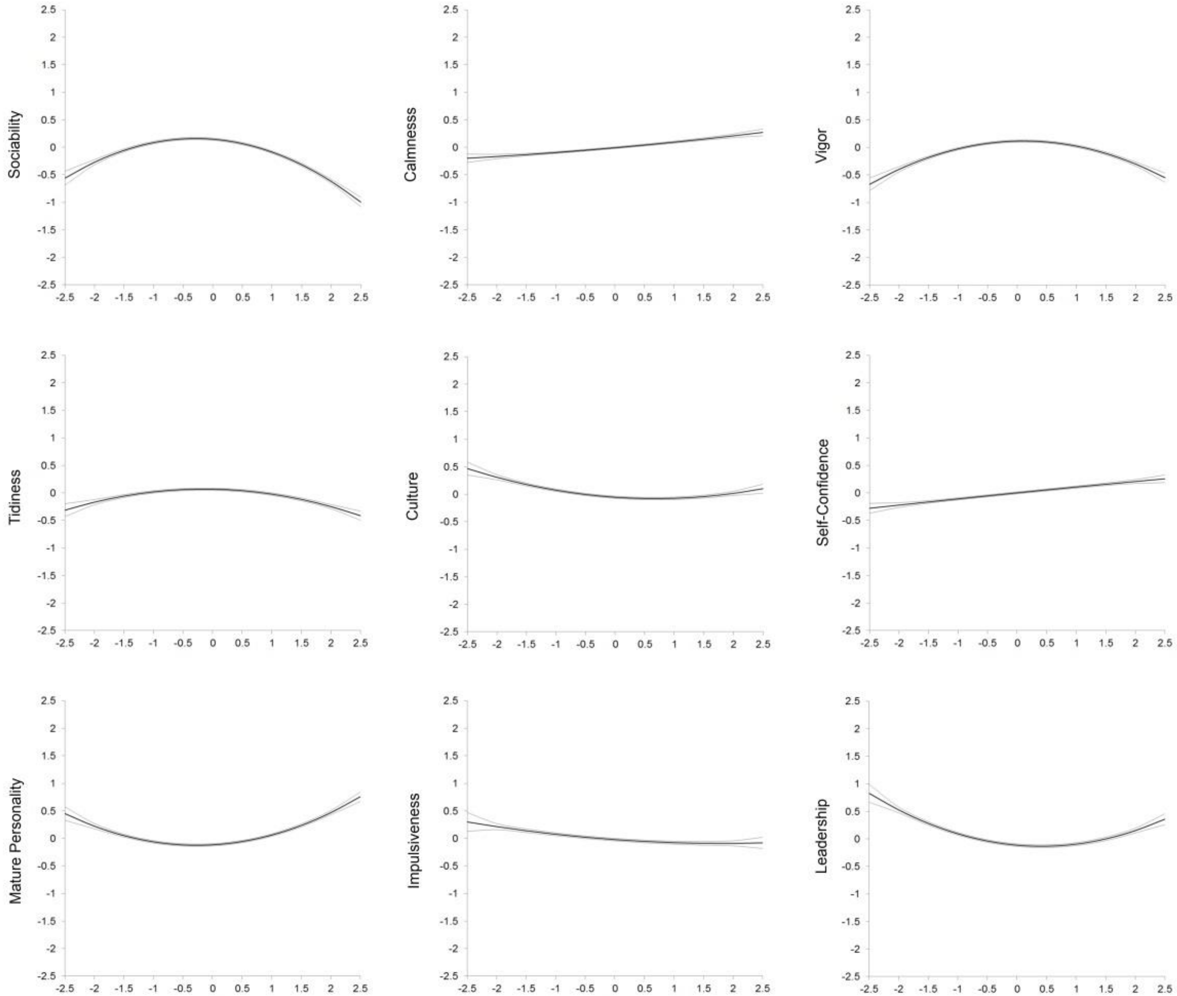
Sociability
Calmnesss
Vigor
Tidiness
Culture
Self-Confidence
Mature Personality
Impulsiveness
Leadership

Figure 4.1

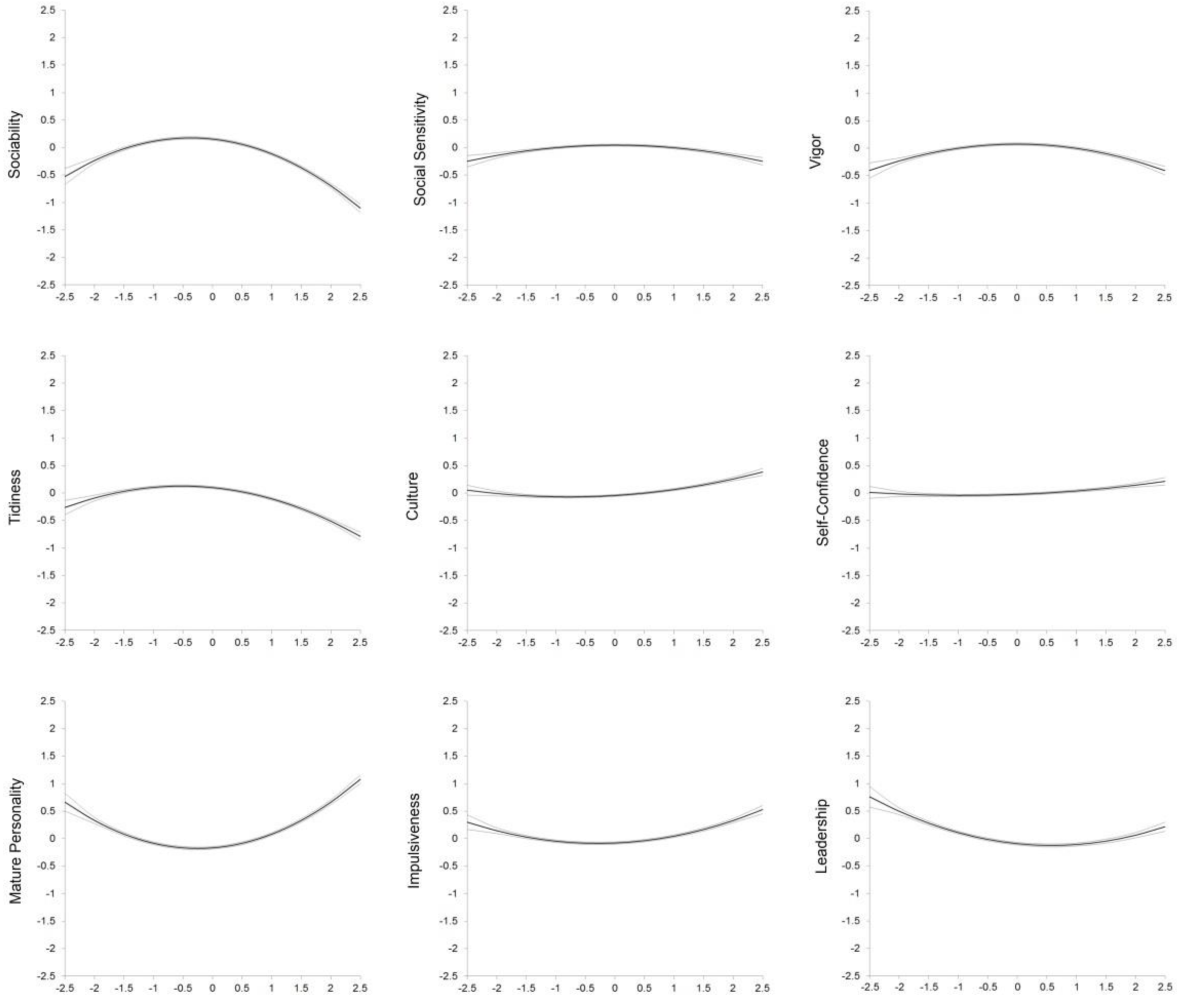
Sociability
Social Sensitivity
Vigor
Tidiness
Culture
Self-Confidence
Mature Personality
Impulsiveness
Leadership
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

Figure 4.2

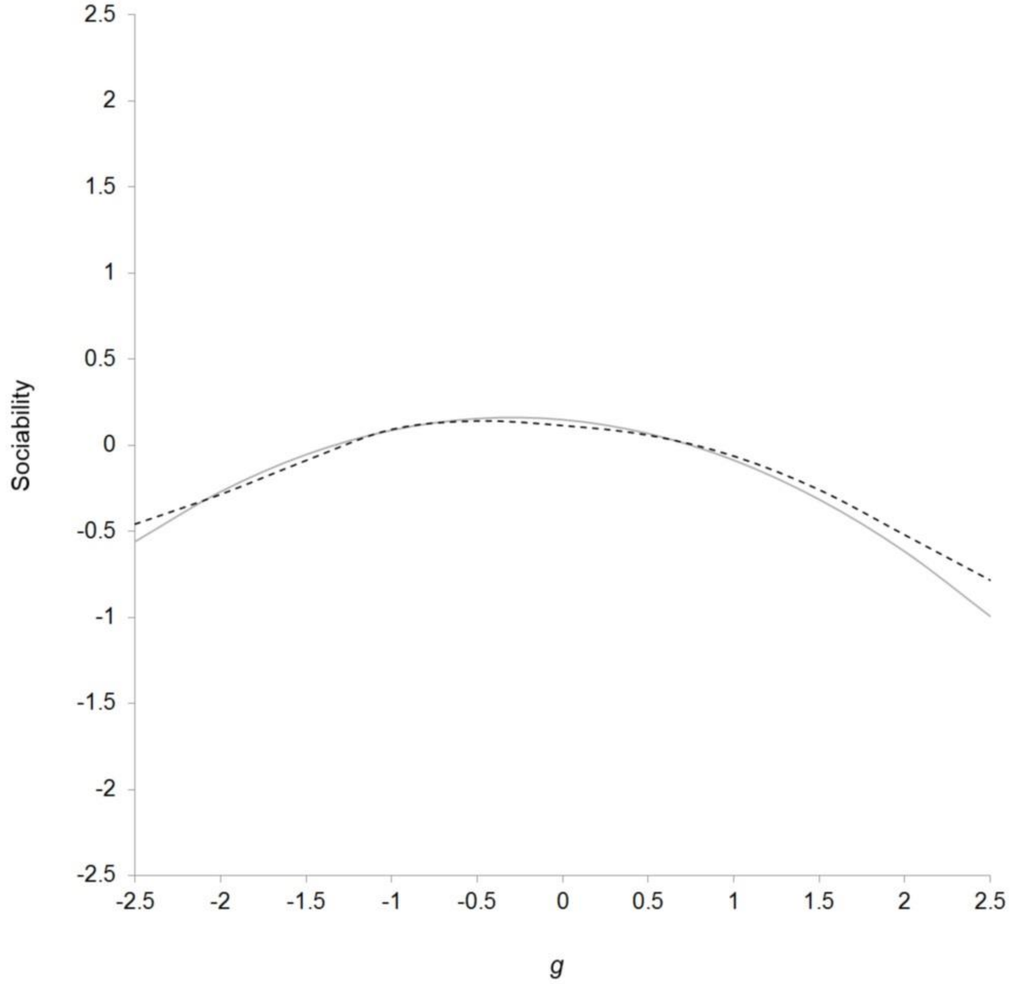

Figure 4.3

4.4 Discussion

In this study we examined linear and quadratic associations between *g* and personality in Project TALENT. SEM was used to estimate linear and quadratic effects of latent *g* on ten personality scales, and the influence of the general factor of personality was controlled by residualizing the personality scores for the GFP. A review of literature provided us with seventeen hypotheses of linear and quadratic associations; nine of these hypotheses (53%) received support in all male samples and ten (58%) received support in all female samples. In this section, we first review the observed associations and discuss in greater detail some of the unexpected and theoretically-relevant results. We then outline limitations of the study, and the implications of our results for future research.

In divergence from previous studies (E. J. Austin et al., 1997; E. J. Austin et al., 2002; Reeve et al., 2006) that have not done so, we found significant quadratic associations of *g* with aspects of personality. Sociability, Vigor, Mature Personality and Leadership were associated in this manner in males, and Sociability, Tidiness and Mature Personality in females. These associations accounted for at most 3.9% of the variance, so it would not be appropriate to conclude that prior studies have misled the field in finding only small quadratic associations. Still, the associations we found would have importance in considering mean personality scores of groups differing greatly from average *g*. In our strongest example, using the grade 10 female sample, negative linear and quadratic associations with *g* predicted a mean Sociability level .70 SDs lower (SE = .02) for individuals two SDs above the mean on *g*, compared to individuals of average *g*[12]. Such a difference would generally be considered substantive, though it did not render the mean Sociability level of 10th grade females with high *g* particularly unusual (the mean fell at approximately the 27th percentile of Sociability of the full sample). The group difference due to the linear effect alone would be only .23 SD (SE = .02), accounting for 1.3% of the variance, compared to 3.7% of the variance in the model with the quadratic effect. This illustrates that failing to consider nonlinear relations causes underestimation of the true associations between certain personality traits and intelligence. Recognition of these nonlinear

[12] For individuals with low *g* (two SDs below the mean), Sociability was .24 SDs below the mean.

associations can be particularly important when focus is on personality in groups with extremely low or high levels of *g* and/or in understanding how the development of personality and intelligence is intertwined in individuals.

Our results were generally consistent with previous findings on intelligence-personality associations in samples of the general population and gifted adolescents (Ackerman & Heggestad, 1997; Zeidner and Shani-Zinovich, 2011). We found that males and females with higher *g* tended to have higher Self-Confidence, and males with higher *g* also averaged higher Calmness; these scales reflected lower Neuroticism in Five-Factor terms (Reeve et al., 2006). For the scales likely reflecting Extraversion, Project TALENT participants with higher *g* scores tended to display lower Sociability but higher Leadership, which was in line with the hypothesis of Ackerman & Wolf (2005) that intelligence is linked to lower social closeness but higher social potency. We found some indirect support for lower Conscientiousness among more intelligent adolescents (for Tidiness, but not Mature Personality), as observed by DeYoung (2011), and lower Agreeableness (Social Sensitivity, but in girls only), as found in gifted studies (Sak, 2004; Zeidner and Shani-Zinovich, 2011). Openness to Experience was incompletely represented by the Culture and Leadership scales, but participants with higher *g* scores were above-average on these scales, with the exception of Culture in grade 9 males. This replicated the most common association in personality-intelligence studies (Ackerman & Heggestad, 1997; DeYoung, 2011).

General intelligence was associated with mean-level differences in all Big Five domains, which is somewhat at odds with existing theories of personality-intelligence relations. For example, Chamorro-Premuzic and Furnham (2006) maintain that each of the Big Five should be related to intellectual competence, but regard Agreeableness as a marginal indicator, and view Neuroticism as mainly being related to intelligence through test anxiety, and Extraversion related through test-taking style. The opposing associations with *g* that we observed for Sociability and Leadership (two aspects of Extraversion), cannot be well-explained within their framework. PPIK theory also does not provide a full explanation for broad associations between *g* and the Big Five. In PPIK theory, *g* is mainly associated with personality due to the involvement of group abilities (such as crystallized intelligence and perceptual speed) in particular trait complexes. Most notably, crystallized intelligence is thought to contribute to the Intellectual/Cultural trait complex, along with Openness to Experience (Ackerman & Heggestad, 1997; Ackerman & Beier, 2006). However, many of the associations we observed in the current study would not be predicted in this framework, such as the association between higher *g* and lower scores on scales reflecting Conscientiousness and Agreeableness, as well as the differential associations of *g* with social closeness and social potency (Ackerman, 2005). For instance, Ackerman & Heggestad (1996) stated that: "Intelligence-as-process correlates weakly with most broad personality factors, except [negatively] for those that are associated with psychopathology" (p. 239).

Our results suggest instead that there are meaningful associations between *g* and each of the Big Five (and/or their facets). Moreover, *g* is closely related to intelligence-as-process or fluid intelligence (Gustafsson, 2002; Kvist & Gustafsson, 2008; Major, Johnson, & Deary, 2012). Overall, our results were more consistent with the personality differences observed in studies of gifted adolescents, such as a greater tendency towards Perceiving (which is correlated with lower Conscientiousness; McCrae & Costa, 1989) and Introversion on the Myers-Briggs Type Indicator (Sak, 2004) and lower Agreeableness in the Big Five (Zeidner and Shani-Zinovich, 2011). Because personality differences have been more apparent in these studies, it seems likely that considering the developmental differences between gifted and normally-developing children and adolescents may be a good way to develop theories of personality-intelligence associations, in addition to examining associations in the general population. It is possible, however, that gifted people may have distinct life experiences (such as experience with accelerated education programs) that make comparisons with general samples more difficult.

Due to quadratic associations of *g* with personality, adolescents with low *g* did not necessarily display the converses of the personality associations of those with high *g*, and in fact were more similar in score with high-*g* students than average ability students on a number of scales. For example, like high-*g* students, they averaged lower Sociability and Tidiness. Participants with low *g* were also unexpectedly found to average higher scores on the Mature Personality and Leadership scales than average-ability students.

Due to the positive association between Mature Personality and Conscientiousness in Reeve et al. (2006), and the negative association of *g* with Conscientiousness in the literature, we predicted a negative quadratic effect of *g* on Maturity. The unexpected positive quadratic effect may reflect the fact that the Mature Personality scale contained several items that tapped self-assessed achievement striving and engagement ("I work fast and get a lot done"; "I am productive"). Thus, it may not be so surprising that students with higher *g* scores (who also tended to have had more success at school) also obtained higher scores on this scale, possibly despite its association with Conscientiousness, on which they tended to score lower.

The most unexpected linear association was the negative association between *g* and the Leadership scale, mostly found in the lower grade levels (in grade 9 males/females, beta = -.12/-.11). This finding may have reflected lack of clear understanding of the items by younger and less able students (e.g. the item "I am influential"). Another possibility is that the students understood the items, but that less intelligent students overestimated their leadership abilities due to lower metacognitive ability to assess their social function (Kruger & Dunning, 1999). This 'Dunning-Kruger' effect may also apply to our finding of higher scores on the Mature Personality scale for individuals with below-average *g*. One final possible interpretation for the Leadership scale finding is that leadership in younger grades is often more social than intellectual in nature, and that social engagement may be negatively related to intellectual performance due to an investment trade-off between social and intellectual activities (Ackerman & Wolf, 2005).

Most of the linear and quadratic associations we observed were present in both sexes. The exceptions to this were linear associations of *g* with Culture, Impulsiveness and Calmness, and a quadratic effect on Social Sensitivity that was only present in females. Based on the content of the Culture scale and its positive association with Big Five Openness (Reeve et al. (2006), we predicted that Culture would show a positive association with *g*. This was the case in females, but small negative associations were found in males (beta = - .067 to -.087). This may reflect the fact that the Culture scale emphasized having good manners over intellectual interests. Perhaps the socialization pressures on girls to be well-mannered were stronger than those on boys. The other sex differences we observed were less readily interpretable.

For scales where an association was found at any grade level, the majority were found in all four grade samples: 13 of 19 in males (68.4%) and 14 of 18 in females (77.8%). This observation supports the view that the effects were not due to chance measurement artifacts from individual samples.

As noted in the Results, even if most effects were consistent, some effects varied substantively in magnitude across grades. These differences have also been the subject of prior theories on personality-intelligence relations (Ackerman & Wolf, 2005; Chamorro-Premuzic & Furnham, 2006). The increase of the negative quadratic effect of *g* on Sociability across grades provides support for the hypothesis that higher social closeness may run counter to the development of intelligence because adolescents with greater need for social closeness select social activities more frequently than (generally) solitary intellectual activities (Ackerman & Wolf, 2005). However, the direction of effect assumed and measured in this study may imply that it was instead higher intelligence that reduced social closeness over time due as higher-*g* students increasingly selecting more solitary activities. Due to the unavailability of item-level data, the PT personality scales were not well-suited to testing the quadratic effect of personality on intelligence (the effect hypothesized by Ackerman & Wolf, 2005). Future studies may be able to disentangle these two effects by comparing the sizes of quadratic effects in each direction.

The negative linear association of *g* with Tidiness became stronger with grade level, which is consistent with the hypothesis of Chamorro-Premuzic and Furnham (2006) that lower intelligence leads to the development of greater orderliness over time as a compensatory mechanism to meet environmental demands[13]. However, the

[13] We tested this hypothesis alternatively by examining differences in mean Tidiness with the sample split by quintile of *g*. Comparing grade 9 to grade 12 samples, Tidiness increased significantly in every quintile ($p < .001$), but there was a progressively greater increase in Tidiness corresponding to lower quintile of *g*.

negative quadratic effect of *g* on Tidiness also indicated that very low levels of *g* corresponded with decreased Tidiness, possibly because very low intelligence is a hindrance to orderly behavior.

There are several limitations surrounding our conclusions regarding personality-intelligence relations. First, our personality scales may not have been measurement invariant across different levels of *g*, which could have caused apparent linear and nonlinear associations that did not exist (McLarnon & Carswell, 2012; Waiyavutti, Johnson, & Deary, 2012). We had, however, no way to test this as we did not have access to the items. Second, we were able to establish that measurement invariance did not hold across samples for *g*. Thus, although we observed some consistency of associations across grades and sexes, the constructs measured across the samples may not have been identical.

The PT personality data had a large GFP that accounted for approximately 40% of the variance in each sample. Our removal of the GFP may have been a limitation because it may have contained substantive personality variance, although most recent research supports a largely artifactual origin of the GFP (Anusic et al., 2009; M.C. Ashton et al., 2009; Bäckström et al., 2009; Chang et al., 2012). One possible explanation for the large GFP in Project TALENT is that the context of in-school testing may have influenced students to "fake good" on the personality scales, and the more intelligent students were more capable and/or more motivated to do so. Given that the main purpose of PT (of which the students were aware) was to assess scholastic talent, it would be most relevant for students to exaggerate scores on scales tapping behaviour socially desirable in the school context (such as diligence and responsibility). The high loading of the Mature Personality scale on the GFP (mean $r = .79$) was consistent with this interpretation, as was the relatively higher loading of Tidiness on the GFP in PT samples compared with the college sample of Reeve et al. (2006). A possible non-artifactual explanation is that more intelligent students were in fact more successfully socialized within the high school environment, and that this led to higher scores on all the PT personality scales. Discounting this, however, studies of students selected for high intelligence found that they did not score higher than unselected students on Agreeableness and

Conscientiousness—Big Five factors that reflect greater socialization (Sak, 2004; Zeidner & Shani-Zinovich, 2011). The alternative of including the GFP could have led to exaggerated *g*-personality associations.

Controlling the GFP in the current study caused a number of the personality scales to have negative linear associations with g, in contrast with the results of Reeve et al. (2006), who found only positive linear associations. Nonetheless, out of eight positive-direction associations in Reeve et al. (2006), six were also found here. The exceptions were the positive associations of *g* with Social Sensitivity and Calmness in females. Replicated associations were with Mature Personality, Calmness and Self-Confidence in males; Mature Personality, Culture, and Self-Confidence in females. These results confirm that the GFP was not entirely responsible for the positive associations observed in the previous study.

Nonetheless, one key implication of our results is that the GFP can be a potentially important confounder or mediator of personality-*g* associations, particularly for linear associations as these relations were the most affected by removal of the GFP (see Supplemental Tables B1 and B2). If the GFP represented at least partly substantive variance instead of methodological variance, it could have been a mediator, whereby the effect of *g* on personality occurred indirectly through the GFP, or vice-versa.

One final limitation to our study was that the sample was assessed in 1960, and relations between personality and intelligence may have shifted since then. This kind of change was observed by Wolf and Ackerman (2005), who that the relation between Extraversion and intelligence was slightly positive before 2000, but slightly negative after 2000. One notable source of such change concerns the erosion of gendered occupational roles since then. Girls at that time had less opportunity to aspire to high education, and especially to occupational achievement in their own names. Moreover, it was very common that they aspired and expected to marry and be supported financially by their husbands. Despite this possibility, females had higher scores than males on the Mature Personality scale at all grade levels, and the association between *g* and Maturity was higher in females than males (see Tables 4.4 and 4.5). Offsetting the age of the sample, one of the strengths of the Project

TALENT sample was that it was representative of the United States in 1960 (Flanagan et al., 1962), so that our results can be generalized to the whole population at that time.

4.4.1 Conclusions and future directions

We found that mean levels for most Project TALENT personality scale scores varied substantially across levels of *g*, and a number of scales showed quadratic associations. These results provide further support for the view that personality-intelligence associations are substantive and relevant to understanding the development of individual differences in both domains (Ackerman, 1996; DeYoung, 2011; Sak, 2004). Our results also indicated two directions for future research in this area: the interpretation of the general factor of personality, and the use of nonlinear models to test the direction of effect (personality on intelligence, or intelligence on personality).

If it was not controlled, the GFP would have had a substantial effect on personality-intelligence relations in the current study due to its positive association with *g*. Future research should examine whether this relation is substantive or artifactual in nature, possibly through the use of multiple raters or social-desirability scales. If the GFP itself is found to be largely artifactual, as much recent research suggests, then it is questionable whether the *g*-GFP association can represent meaningful variance, but research in this area is still ongoing.

The potential to examine direction of effect deserves more consideration in personality-intelligence research. In the current study, we focused on the associations of the quadratic function of *g* with personality, but such nonlinear associations may be found in the other direction, or in both directions. Although the nonlinear associations we observed were small in terms of variance explained, they were capable of resulting in substantive personality differences for individuals at the extreme ends of the *g* distribution. Nonlinear associations that result in substantive differences in personality at the tails of the intelligence distribution, or differences in intelligence at the tails of personality distribution, can potentially be very informative about how personality and intelligence interact with each other. In spite of this, it is likely that the direction of influence runs both ways in most cases, and that the strength and direction varies depending on the environment and over time. In order to understand the interplay between personality and intelligence more complex study designs and models are needed.

4.5 Integrating cognitive abilities and interests

After examining personality-intelligence associations in PT, the next step was to bring occupational interests into the research. As presented in the introduction, I decided to focus on the possibility of replicating the trait complexes composed of cognitive abilities and interests in PPIK theory (Ackerman & Heggestad, 1997). The proposed capacity of these trait complexes to predict future occupation is a key aspect of the theory, and a strength of the PT dataset was its longitudinal data. Thus the trait complexes were examined in the context of occupation eleven years after high school, where the hypothesis was that they should have equal predictive validity to the use of individual scores for cognitive abilities and interests. This was tested for trait complexes composed both of factors and latent classes. In addition, the trait complexes were evaluated in terms of how well they matched their descriptions in PPIK theory.

Chapter 5: Trait complexes of cognitive abilities and interests and their predictive validity for occupation

5.1 Introduction

Intelligence and interests are both important predictors of occupational attainment and job type (De Fruyt & Mervielde, 1999; Kuncel, Hezlett, & Ones, 2004; Schmidt & Hunter, 2004). General intelligence (*g*) relates strongly to occupational level (for review, see Schmidt & Hunter, 2004) Specific abilities such as spatial and verbal abilities are also relevant to employment in specific occupational areas such as the humanities and scientific fields (Johnson & Bouchard, 2009; Wai, Lubinski, & Benbow, 2009). In keeping with their applied purpose, measures of occupational interests are predictive of the nature of future employment (J. T. Austin & Hanisch, 1990; De Fruyt & Mervielde, 1999), as well as of performance in that employment (Van Iddekinge, Putka, & Campbell, 2011).

Given the predictive validity of interests and cognitive abilities, discovering any overlap between them could have theoretical as well as applied significance (Johnson & Bouchard, 2009). Greater understanding of the links between cognitive abilities and interests could aid our theories of both cognitive and interest development. Some researchers have proposed that cognitive abilities have substantial roles in the development of interests. Gottfredson's (1986, 2005) theory of circumscription and compromise posits that individuals' self-awareness of their levels of general intelligence influences their interest in particular occupations according to their cognitive complexity. Hogan and Roberts' (2000) socioanalytic model of identity development proposed that interests are built on successful experiences with cognitive investment, which depend on intelligence. In turn, interests are theorized to influence the development of intelligence through the selection of future learning environments (Hogan & Roberts, 2000; Scarr, 1996). Development of a framework that better integrates interests and cognitive abilities could lead to better career counseling advice, and long-term increases in person-environment fit, which refers to matches between work environments and individuals' abilities and preferences (Dawis & Lofquist, 1984; Holland, 1997).

Increases in person-environment fit could have benefits for both individuals' work satisfaction and productivity.

In spite of the potential importance of integrative research, there have been relatively few studies of the associations between interests and cognitive abilities (Ackerman & Heggestad, 1997; Anthoney & Armstrong, 2010; Johnson & Bouchard, 2009). Moreover, these studies have often been hindered by the use of college samples with restricted ability ranges, and failure to separate specific abilities from general intelligence statistically and conceptually (Johnson & Bouchard, 2009). For example, Ackerman & Heggestad (1997) reported meta-analytic associations of interests with nominally specific abilities, but these ability measures were not statistically independent of *g*, leaving it unclear to what degree the associations were ascribable to *g* or specific abilities. At the same time, general theories of interest-ability interaction such as Hogan's and Gottfredson's have not provided detailed hypotheses on the overlap between specific abilities and interests, predictions which might be the most useful for practical and theoretical reasons. In spite of the limitations of Ackerman and Heggestad's (1997) meta-analysis, PPIK theory (intelligence-as-process, personality, interests, and intelligence-as-knowledge) by Ackerman (1996) did address this specific overlap.

Ackerman and colleagues proposed that intelligence, personality and interests coalesce into four "trait complexes" (Social, Clerical/Conventional, Science/Math, and Intellectual/Cultural) (Ackerman, 1996; Ackerman & Heggestad, 1997). They defined trait complexes as similar to Snow's (1963) aptitude complexes, or "combinations of levels of some variables which are particularly appropriate for efficient learning" (p. 120, cited in Ackerman & Beier, 2003a); however, Ackerman and colleagues focused on attainment of academic knowledge/expertise and practice of particular occupations rather than the learning processes necessary to reach those states. Like Hogan & Roberts (2000), they regarded the interaction between interests and cognitive ability as reciprocal. They proposed that particular abilities and interests become more strongly related throughout development because the abilities are suited to success in certain domains, and the satisfaction brought by this success spurs the interest that motivates further cognitive investment. Trait complexes are thus held to influence subsequently how individuals select and attain particular occupations through their roles in knowledge acquisition in academic/occupational settings (Ackerman, 1996).

In PPIK theory, interests are conceptualized in terms of the RIASEC model of occupational types by Holland (1973, 1997; Realistic, Investigative, Artistic, Social, Enterprising and Conventional). The RIASEC model is the predominant model of occupational interests in the literature. It is based on preferences for six types of work environments that are organized in a hexagonal circumplex, with adjacent types more closely related in job demands than opposite types. Research on the RIASEC hexagon has supported this structure for interests and employment types in the United States (Holland, 1997; Tracey & Rounds, 1993) A significant limitation is that the model does not address the roles of cognitive abilities in interests. However, given the consistent associations between RIASEC interests and cognitive ability measures that Ackerman and Heggestad (1997) found, they used its framework to formulate PPIK theory (Ackerman, 1996). The RIASEC model is well-suited to an integrative framework because work environments are important (though not the only) contexts in which ability and non-ability traits converge (Armstrong et al., 2008).

PPIK theory also specified that only two of the four trait complexes primarily involve ability-interest associations; the other two primarily involve personality-interest associations (Ackerman, 1997). Based on a meta-analysis of five studies, Ackerman and Heggestad (1997) observed that only three of the six Holland types were related substantially to cognitive abilities: Realistic interests, defined as interests in activities involving physical action and motor coordination; Investigative interests, defined as interests in cognitive problem solving; and Artistic interests, defined as interests in expression through artistic media (Holland, 1973). Realistic and Investigative interests were associated, and each was also associated with general intelligence (intelligence-as-process in PPIK theory), as well as with math and spatial abilities, relations which formed a Science/Math trait complex (for example, the meta-analytic correlation of spatial ability with Realistic interests was .28; Figure 5; Ackerman, 1996). Artistic interests were held to be associated with crystallized or verbal ability (intelligence-as-knowledge in PPIK theory), forming an Intellectual/Cultural trait complex (e.g. a correlation of .36 between Artistic interests and verbal ability; Ackerman, 1996). The other two trait complexes (Social and Clerical/Conventional) showed moderate associations with personality, but only minor and/or less consistent associations with cognitive abilities.[14] While personality is likely to be important in understanding occupational outcomes in certain domains, it may be helpful to further the understanding of interest-ability associations exclusive of personality before attempting to integrate all three sources of individual differences. The latter was the focus of the current study. As we did not include personality in our analyses, we anticipated finding only the Science/Math and Intellectual/Cultural trait complexes.

Ackerman and colleagues found that their trait complexes were moderately to strongly related to academic knowledge (Ackerman & Rolfus, 1999), selected university course (Ackerman, 2000), and university course performance (Kanfer et al., 2010), which are outcomes along the path to vocational choice. However, trait complexes in these studies were obtained through factor analysis, which may be inconsistent with how trait complexes have been conceptualized. Factor analysis is used to group variables, and relies on the assumption that the groupings apply in the same way to all individuals in the population. A method of analysis that is arguably better suited to identifying trait complexes is latent class analysis (LCA). LCA groups individuals together based on their scores of a set of variables, ignoring the associations among the variables at the population level. Thus, latent classes can represent groups of individuals who have “combinations of levels of some variables” (Snow, 1963), rather than sets of positions on groups of variables.

One study that used LCA to define interests groups was conducted by Johnson and Bouchard (2009). They examined the mean-level differences in cognitive abilities among eight latent interest classes, where cognitive ability was defined according to an updated version of Vernon’s intelligence model, the Verbal,

[14] The Conventional trait complex involved a positive correlation between Conventional interests and perceptual speed (e.g. .15 in Rolfus & Ackerman, 1996), and the Social trait complex small negative correlations between Social interests with math ability (-.21 in Rolfus & Ackerman, 1996) and spatial ability (-.13 in Randahl, 1991).

Perceptual and Image Rotation (VPR) model (Johnson & Bouchard, 2005b). Based on previous research in the same sample, Johnson & Bouchard (2009) separated intelligence into orthogonal factors for general intelligence (*g*) and two residual dimensions: Verbal-Image Rotation and Focus-Diffusion (Johnson, Bouchard, et al., 2007). Mean levels of *g* varied strongly among the eight latent classes. Beyond this, however, latent classes of interests in leadership, exploration and adventure were related to Image Rotation abilities, whereas interests in cultural and persuasion occupations were related to Verbal abilities. These results were consistent with Vernon's (1961) conceptualization of interests in his verbal-perceptual model of intelligence. Vernon proposed that verbal and math abilities were related to achievement and interest in traditional educational (math and verbal subjects). On the other hand, perceptual (spatial and mechanical) abilities were related to aptitude for technical, scientific and practical subjects. Although Ackerman and Heggestad (1997) did not statistically isolate g from specific abilities, these two themes were apparent in their Math/Science and Intellectual/Cultural trait complexes. In addition, Johnson and Bouchard's (2009) results supported the important role of *g* in occupational interests (Gottfredson, 1986, 2005), such that mean levels of intelligence varied strongly across the interest classes, being highest for Science and lowest for Personal Care. This role for *g* is not as central in the trait complexes of PPIK theory.

Johnson and Bouchard's (2009) study was not intended to test the concept of trait complexes. In order to examine whether there are trait complexes of interests and abilities, both variables should be entered simultaneously into latent class analysis, so that groups with different levels of interests and abilities may be identified. These groups would then reflect the integration of interests and abilities in interlocked transactions implied by Ackerman and Beier's (2003) definition of trait "trait complex". The main purpose of the current study was to test the validity of the trait complex concept. This was done by comparing the ability of trait complexes to predict occupational type with the individual scale scores for cognitive ability and interests. We hypothesized that if trait complexes are true groupings of individuals which influence the likelihood of acquiring specialized occupational knowledge, and thus influence career choices (Ackerman & Beier, 2003a), then latent classes representing them should show strong predictive validity for type of future employment. Ackerman and Beier (2003) put this question similarly: "is there a synergy among elements within the trait complexes, so that concentrating on trait complexes is more informative in the career choice context than individual trait measures?" (p. 209). Due to the prominent position that trait complexes occupy in PPIK theory, we predicted that they should demonstrate predictive power (in terms of explained variance) at least equal to the use of individual scores for cognitive abilities and interests. Equality would be accepted as being in favour of trait complexes because of the greater parsimony of latent classes. As a point of comparison with the research of Ackerman and colleagues, we also compared the predictive validity of factor-analytic trait complexes to the individual scale scores, with the same predictions for these trait complexes as for the LCA trait complexes.

The current study was thus meant to address two questions: first, would exploratorily-derived trait complexes of interests and abilities, obtained through LCA and factor analysis, replicate the content of the Science/Math and Intellectual/Cultural trait complexes proposed by Ackerman and Heggestad (1997)? The Social and Clerical/Conventional trait complexes were not anticipated due to the exclusion of personality variables in the present analysis. Second, would the trait complexes obtained by either method display predictive validities at least equal to those of individual scale scores? The predictions of PPIK theory would be contradicted if the trait complexes obtained did not resemble the Science/Math and Intellectual/Cultural trait complexes, and/or if the trait complexes did not display predictive validities comparable to those of individual scale scores. Additionally, if only the trait complexes derived by factor analysis satisfied these conditions then it would raise theoretical questions about their definition as combinations of levels of the variables.

5.1.2 Previous Project TALENT research

For this study, we made use of data from Project TALENT (PT). PT was a longitudinal study of American high school students meant to investigate their aptitudes, interests, and backgrounds, and the influences of these variables on educational and occupational outcomes (Flanagan et al., 1962). During the study, 60 aptitude tests and extensive interest scales were given to a sample of over 400,000 American high school students, who were representative of the U.S. student population. The students were followed up over an 11-year period after high school and surveyed on their education and occupational experiences.

Two previous have studies have used PT intelligence and interest data to predict occupational type (Austin & Hanisch, 1990; Humphreys, Lubinski & Yao, 1993). Austin and Hanisch (1990) examined the tenth-grade PT sample and found five discriminant functions that predicted the 12 occupation categories defined by PT investigators. The results indicated that occupational category could be predicted above chance for 10 of the 12 categories (exceptions were Technical and Sales jobs). Two major discriminant functions described the interest and ability data. The first discriminant function, interpreted as verbally-oriented general mental ability, mainly predicted occupational prestige or level. The second function, which differentiated individuals based on mathematics, spatial ability and gender, predicted scientific and technical occupations (Hanisch & Austin, 1990). These functions, however, were not interpreted in a trait-complex framework, and their capacities to predict occupation were not compared to those of individual scales.

Humphreys et al. (1993) explored the differential prediction of occupation for groups defined by the top 20 percentiles of spatial and verbal abilities. They attempted to equalize these groups for general intelligence by selecting students in the top 20 percentiles on composites of spatial-math and verbal-math scores, math ability being used as a proxy for general cognitive ability. They found that the high-verbal and high-spatial groups had significantly different probabilities of entering scientific/engineering and humanities jobs. The groups also differed strongly in their mean occupational interests, with the high spatial-ability group showing greater interest in mechanical-technical jobs, and the high verbal-ability group more interested in literary-linguistic jobs. Although these groups could be considered similar to trait complex groups, they were pre-specified and not derived through any empirical analysis. In addition, the composite method of Humphreys et al. (1993) was not entirely successful in controlling for *g*, as the high verbal group had higher mean scores for math than the high spatial group (see Table 3 of Humphreys et al., 1993). As in other studies on ability-interest associations, a significant limitation of both Austin and Hanisch (1990) and Humphreys et al. (1993) was the lack of independence of the specific ability measures from general intelligence. In summary, the current study tested the trait complexes of PPIK theory in two novel ways, by employing latent class analysis to define trait complexes, and by comparing their predictive validity for future occupation with individual scale scores for cognitive abilities and interests.

5.2 Method

5.2.1 Sample

The two highest-grade samples of PT were used (grades 11 and 12), because we considered the older students more likely to have considered their future career prospects. The use of two samples allowed for replication of potential trait complexes across samples. In grade 11, there were 47,027 females and 45,292 males. In grade 12, there were 41,456 females and 39,674 males. The total sample size was 173,449, with 51.0% females. The mean age was 16.4 in grade 11 (SD = .69) and 17.3 in grade 12 (SD = .67). Males were slightly older than females in both samples: 16.4 compared to 16.3 in grade 11, and 17.4 compared to 17.2 in grade 12. The full individual age range was 8 to 21, the younger participants having skipped multiple grades, and the older participants having been held back.

For the 11-year follow-up mail survey, responses were obtained from 27.5% of the original grade-11 sample, and 30.9% of the grade-12 sample. In order to adjust for the lack of representativeness of the follow-up sample, PT investigators conducted special interviews with non-respondents to the mail questionnaires. Approximately 2500 participants in each grade cohort were given telephone or in-person interviews (Wise et al., 1979). Sample weights were created in accordance with the sampling ratio of the special sample to original the 1960 sample (Wise et al., 1979). We applied these sample weights to our analyses where follow-up occupation data were used. The follow-ups were conducted in 1971 for the grade-12 sample and 1972 for the grade-11 sample. Participants were not asked their ages at follow-up; however the dates of the follow-up surveys were recorded. The mean week of the year that the surveys were received was week 7.1 in the grade-12 sample (SD = 6.3) and week 8.2 for the grade-11 sample (SD = 8.0). As the baseline data were collected in March 1960, it can be inferred that the mean follow-up sample age was approximately ten years and eleven months older than baseline for the grade-12 sample. For the grade-11 sample it was eleven years and eleven months older than at baseline. Thus, their mean ages were approximately 28.3 for the grade-11 sample and 28.2 for the grade-12 sample. The follow-up sample was 52% female according to gender recorded at baseline.

5.2.2 Intelligence measures

The intelligence measures were chosen from the 60 aptitude and achievement tests in PT (Wise et al., 1979). To ensure that our measures of specific ability did not rely on overly specialized knowledge, we excluded the "information tests" on academic and non-academic topics. Our starting point was the 22-test narrow selection as defined in our previous study (Major et al., 2012; see descriptions of the tests and reliabilities there). However, we excluded three of the English achievement tests (Spelling, Capitalization and Punctuation) because they relied too heavily on knowledge acquired in school classes. The Vocabulary test was also excluded because it was part of the original information tests. Eighteen tests remained after this selection. In previous studies, the advanced math test was omitted because it was designed for students above the tenth grade, and was thus deemed unfair for younger students. However, it was also excluded here because we initially planned to use all four grade samples.

Data screening was the same as in Major et al. (2012). Scores on the PT response credibility index, based on a screening test assessing illiteracy, mental disability or apathetic testing attitude, were used to exclude participants who did not reach the cut-offs set by the PT study designers, except where only mental slowness was indicated (a low score for the number of responses on the Clerical Checking test). Transformations were applied to two tests that displayed non-normal distributions (English Usage and Table Reading): English Usage was negatively skewed, and we applied a square-root transformation (the direction of the variable was reversed prior to transformation and then reversed back). Table Reading was strongly positively skewed and leptokurtic and thus had logarithmic and cosine transformations applied. Cases showing severe problems with multivariate outliers were also removed (Major et al., 2012). Following data screening of the intelligence tests, total sample size was reduced to 170,723 (1.57% of the total sample removed, the majority for failure on the response credibility index). After removal of these cases, some missing data remained for each cognitive ability test. In the male samples, 2.1 to 2.8% of scores were missing, while 2.1 to 2.7% were missing in the female samples.

5.2.3 Interest measures

The PT interest scales consist of 17 composites that were designed to capture interests in different job areas, such as Artistic and Mechanical-Technical jobs (Wise et al., 1979). However, a limitation of these scales is that they were created on an a priori basis, without regard to the observed correlations among the items, or to any particular theoretical framework. Because item-level data were available, we derived new interest scales based on exploratory factor analysis. Previous studies have employed the pre-existing scales (J. T. Austin & Hanisch, 1990; Humphreys, Lubinski, & Yao, 1993). It was anticipated that these new scales would have greater validity and thus predictive power for occupation than the original PT interest scales

The original interest scales were formed from 205 items, 122 of which were occupation titles (e.g. musician, rancher, etc.) and 83 of which were activities applicable to work and school settings (e.g. typewriting, selling furniture). Students were asked to indicate how well they would like or dislike the occupation or activity, and were instructed to disregard educational requirements, salary, social standing, or other factors (Wise et al., 1979). Responses were recorded on a 5-point Likert scale from “I would dislike this very much” to “I would like this very much”.

For the occupation titles, missing data percentages for the items for each sample ranged from 2.6-6.7%, and were very consistent across grade and sex. An exception was 21 consecutive items in each sample that had greater numbers of omitted responses, likely due to a coding error in the PT database. For example, in grade-12 males, these 21 items had total missing data percentages of 7.9-9.0%. Because these additional missing data were missing at random, they should not have biased the analyses. Three occupation items were excluded from the male samples because they were female-oriented occupations that received very low endorsements (Maid, Dish Washer and Housewife).

5.2.4 Occupational categories

Table 5.1 contains the occupation category titles and sample percentages for grade-12 males and females with 11-year follow-up data. Participants were asked to state their current job titles. These written responses were reduced to 254 job codes representing specific jobs or job areas such as Airplane Navigator, Veterinarian or Metal Trades (Wise et al., 1979). These job titles were then organized by PT investigators into twelve categories according to broad occupational themes. The most prevalent job category in males was Business Administration, while in females it was Clerical and Office Work. Due to the period during which the data were collected, there were large gender differences in the frequency of different occupational groups.

Table 5.1
Occupation categories and sample percentages (grade 12 sample).

Category title	Males (%)	Females (%)
Physical Sciences, Engineering and Mathematics	5.5	0.2
Medical and Biological Science	2.5	3.1
Business Administration	19.3	3.1
Teaching and Social Service	8.1	8.3
Humanities, Law, and Social Science	3.4	0.9
Fine and Performing Arts	1.0	0.4
Technical	5.3	1.9
Sales	11.2	1.9
Mechanical and Industrial Trades	8.9	0.5
Construction	7.8	0.02
Clerical and Office Work	3.4	14.7
General Labour and Public Service	15.2	7.6
Vague and Undesignated	8.3	7.9
Housewife	N/A	49.5

5.2.5 Method of analysis

The analysis was done in three steps. First, factor analysis was performed separately on the interest items and cognitive ability scales, and the factor scores were retained. Second, trait complexes were obtained from the interest and cognitive ability scores through both factor analysis and latent class analysis, conducted in Mplus 5.21. In the third step, the trait complex data (class membership and factor scores) were used to predict future occupational category using logistic regression in Mplus and multinomial regression in SPSS 18. As the first step was preliminary to the two main analyses, its results are presented in this section.

Missing data were handled through maximum likelihood estimation, which assumes that the data were missing at random (MAR). This assumption was tenable because it is unlikely that students purposely avoided particular cognitive ability or interest tests.

5.2.6 Interest and cognitive ability factors

The 205 items were allocated to 17 original interest scales by PT designers on an atheoretical basis. We derived new scales based on exploratory factor analysis (EFA). Two separate EFAs were conducted for the occupation titles and activities. The analysis was conducted in SPSS with Promax rotation (kappa = 3).

Examining the scree plots in the grade-11 and -12 male samples suggested seven factors, while it was less clear in the female samples. When further factors were extracted beyond seven in males, these also displayed adequate simple structure and interpretability up to the tenth factor. Upon extracting ten factors in females, nine of these were recognizable counterparts to the male factors, except the tenth, which was not easily interpretable and obtained no loadings above .35 in either the grades-11 or -12 samples. Thus, ten interest factors were retained in both males and females.

The names assigned to the factors in grade-12 males and their two highest factor loadings were as follows: Trades (Riveter: .77, Bricklayer: .73), Politics (U.S. Congressman: .98, U.S. Senator: .96), Science (Chemical Engineer: .78, Electrical Engineer: .76), Business Clerical (Bookkeeper: .75, Office Clerk: .69), Arts (Artist: .87, Writer: .73), Military (Air Force Officer: .77, Army Officer: .74), Teaching (High School Teacher: .92, School Principal: ..78), Medical (Doctor: .80, Surgeon: .79), Business Sales (Stock Salesman: .55, Insurance Agent: .55), and Architecture (Designer: .49, Architect: .48). The total variance explained by the ten factors was 45.1%.

The factors in grade-12 females were assigned the same names as the males; the following were the two highest loadings in the grade-12 sample: Trades (House Painter: .72, Deliveryman: .70), Politics (U.S. Congressman: .96, U.S. Senator: .93) Science (Electrical Engineer: .74, Chemical Engineer: .67), Business Clerical (Typist: .77, Secretary: .75), Military (Air Force Officer: .83, Marine Corps Officer: .80), Teaching (High School Teacher: .70, School Principal: .60), Medical (Nurse: .70, Doctor: .70), Business Sales (Insurance Agent: .55, Personnel Administrator: .55). The total variance explained by the nine factors was 41.8%.

Factor analysis was also performed on the 83 activity items. Judging by scree plots and interpretability, six factors were found in the grade-12 male sample. However, upon examining their contents and correlations with the occupation title factors, it was found that five out of six of these factors were redundant with the occupation title factors. For example, there was a factor composed of activities relevant to trades occupations (e.g. “repair an auto”, “work in a steel mill”), which correlated highly with the trades factor for the occupation titles (r = 0.76). Activities factors were found with moderate to high correlations with the occupation factors that were labelled Science, Business Clerical, Arts, and Business Sales (mean correlation = .65, SD = .12). The remaining activity factor was composed of Sports activities, and correlated most highly with the Military occupation factor (r = .35). Because of the greater robustness of the occupation title factors (due to the greater number of items), and the redundancy of the activities factors, we decided to retain only the occupation title factors to represent the occupational interests. The activity factors were not used.

Due to the presence of missing data, the factor scores for the interest factors could not be obtained without the exclusion of participants with incomplete data. Thus, we constructed composites using the uniformly-weighted means of the non-missing item scores, selecting items that loaded .30 or above on the respective factors. The composites correlated highly with the factor scores for participants with complete data. Excluding the Architecture composite, in grade-12 males, the composites had correlations with their factor scores that ranged from .92 to .99. The correlation of Architecture composite with its factor was lower (.53). This was due to the lower factor loadings for the Architecture factor (see above). In grade-12 females, the composites had correlations with the factor scores that ranged from .93 to .99. The range of correlations was similar in the grade-11 samples: in grade-11 males it was .78 to .99 (Architecture: .45), in grade-11 females it was .94 to .99.

As in Johnson & Bouchard (2007), we sought to obtain specific cognitive ability scores that were separate from general intelligence. To do this, we extracted the general factor of the 18 tests using maximum likelihood estimation and obtained *g* factor scores. The *g* factor explained a mean of 35.6% of the variance in the male samples, and 36.4% in the female samples. We regressed the individual test scores onto the *g*-factor scores, and entered the residuals into EFA in Mplus. In all four samples, the scree plots suggested four residual factors and the 4-factor EFA solution displayed good fit statistics. For example, in grade-12 males these were: RMSEA, .043 (90% confidence interval: .042 - .044), SRMR, .026; in grade-12 females: RMSEA: .042 (.041- .043), SRMR: .018. The four residual factors were labelled Spatial (made up of tests requiring spatial reasoning), English (loadings from the English tests and Memory for Words), Speed (loadings from all the speeded tests), and Math (a bipolar factor on which the Math tests and Arithmetic loaded positively, and three Verbal tests loaded negatively).

Based on these exploratory results, we created a confirmatory bi-factor model in each sample. In the bi-factor model, *g* is allowed to influence each test score, and specific abilities form their own factors that are uncorrelated with *g*. Factor loadings on the four factors were specified if they were .15 or greater in the EFA results. Table 2 displays the factor loadings for the confirmatory bi-factor models in the grade-12 samples. In the grade-12 males the specific ability factors accounted for 2.0% to 9.1% of the variance, in the grade-12 females they accounted for 1.5% to 8.5% of the variance. The same factor model forms were specified in the grade-11 samples, but we allowed the loading parameters to vary freely and did not specifically test for measurement invariance. Model fit was good in all samples. The fit statistics in each sample were as follows: grade-11 males: CFI: .983, RMSEA: .036 (.036 - .037), grade-12 males: CFI: .980, RMSEA: .040 (.040 - .041), grade-11 females: CFI: .984, RMSEA: .034 (.034 - .035), grade-12 females: CFI:

.982, RMSEA: .037 (.036 - .038). Factor scores for *g* and the specific abilities were saved for further analyses.

Table 5.2
Factor loadings for grade 12 males/females in the confirmatory intelligence model.

Test Name	Factor				
	g	Spatial	English	Speed	Math
Memory for sentences	.27/.35				
Memory for words	.49/.55		.13/.09		
Disguised words	.62/.64		.22/.20	.30/.31	-.16/-.15
English usage	.62/.64		.41/.39		
Effective expression	.58/.57		.31/.24		
Word functions in sent.	.73/.77		.13/.06		
Reading comprehension	.84/.87		.11/–		-.18/-.18
Creativity	.71/.68	.17/.17			-.22/-.14
Mechanical reasoning	.62/.62	.53/.44			
Visualization in 2D	.40/.41	.46/.42		.23/.18	
Visualization in 3D	.53/.55	.58/.54			
Abstract reasoning	.67/.68	.31/.28			
Math 1	.79/.77				.17/.19
Math 2	.83/.73				.34/.39
Arithmetic comp.	.56/.55			.34/.31	.32/.31
Table reading	.22/.24			.70/.70	
Clerical checking	.08/.09			.73/.71	
Object inspection	.13/.21	.29/.25		.60/.57	

Note: All freely-estimated factor loadings are shown and significant ($p < .001$)

5.3 Results

5.3.1 Factor-analytic trait complexes

Table 5.3 displays the correlations between the abilities and interests in the grade-12 samples. They reveal that most of the interest composites were positively related to *g*, except those scales relating to non-professional or semi-skilled jobs, which had negative correlations (Trades and Clerical interests in males, and Clerical interest in females). The scores for residual abilities displayed more differentiated, and generally much lower, correlations with interest scales. Two notable correlations were between Spatial ability and Science interest, and between English ability and Arts interest. Although *g* and the residual abilities were orthogonal in the intelligence model, their factor scores had slight non-zero correlations, some of which were as large as those between interests and abilities.[15]

[15] We removed the small amount of remaining *g* variance from the residual ability scores through regression and re-ran the factor and latent class analyses. This did not substantially alter the characters of the trait complexes obtained through either one.

Table 5.3
Correlations matrix for interest composites and cognitive ability factors (grade 12 males/females)

	Trades	Politics	Sci.	Cler.	Med.	Arts	Teach.	Milit.	Sales	Arch.	g	Spatial	Eng.	Speed	Math
I: Trades	–	.28	.49	.25	.23	.32	.23	.46	.37	–	.03	.10	-.06	-.06	.01
I: Politics	.07	–	.44	.12	.31	.50	.49	.41	.57	–	.17	-.06	-.02	.01	.07
I: Science	.26	.30	–	.01	.59	.45	.32	.42	.38	–	.30	.09	-.08	-.04	.14
I: Clerical	.33	.40	.21	–	-.08	.05	.18	.13	.51	–	-.20	.02	-.04	.11	.03
I: Medicine	.02	.40	.47	.24	–	.31	.30	.29	.23	–	.16	.00	-.02	-.02	.08
I: Arts	.14	.47	.34	.35	.40	–	.49	.36	.53	–	.31	.01	.06	-.04	-.04
I: Teaching	.14	.53	.24	.45	.38	.54	–	.27	.50	–	.22	-.09	.03	.00	.13
I: Military	.28	.36	.42	.23	.29	.29	.26	–	.45	–	.10	.01	-.03	-.02	.01
I: Sales	.19	.63	.29	.67	.41	.56	.56	.40	–	–	.10	-.04	-.02	.02	.03
I: Architecture	.33	.34	.49	.34	.31	.63	.31	.35	.45	–	–	–	–	–	–
g	-.26	.14	.25	-.08	.17	.14	.11	.09	.08	.10	–	.06	.07	-.01	.16
Spatial	.12	-.14	.14	-.13	-.07	-.04	-.16	.01	-.18	.13	.08	–	-.42	.12	-.03
English	-.12	.05	-.11	.00	.04	.13	.09	-.01	.09	-.04	.13	-.34	–	-.07	-.12
Speed	-.04	.07	.00	.08	.05	.00	.03	.02	.06	.01	.02	.04	-.06	–	-.08
Math	-.05	.06	.10	.11	.03	-.07	.07	-.03	.04	-.02	.05	-.25	-.21	-.07	–

Note: Females are above the diagonal, males are below

We examined exploratory factor solutions for the interests and cognitive abilities with additional factors one at a time to assess their fit and interpretability, beginning from the one-factor solution. Model fit improved markedly from one to three factors in all samples. Fit also improved from three to four factors, but the four-factor solutions contained significant problems in both males and females. The fourth factor in the grade-12 males was a near-singlet factor with Arts interest loading above 1, and the four-factor solution did not converge in grade-12 females. Thus, the three-factor EFA solution was used as a basis for constructing a CFA model for the trait complexes, in conjunction with modification indices. Tables 5.4 and 5.5 display the standardized factor loadings for the CFA model in males and females, respectively. As the aim of the trait complex model was to capture covariance between interests and cognitive abilities, correlated residuals were allowed if they were within the same domain (i.e. within interests or within cognitive abilities). In the male samples there were positive correlated residuals between Medicine and Science interests, Sales and Clerical interests, Architecture and Arts interests, and Verbal and English residual abilities. There were negative correlated residuals between English and Spatial ability. In the female samples there were only two correlated residuals, positive between Military and Trades interests, and a negative residual between English and Spatial ability. The fit statistics of the models were only very marginally acceptable. They were as follows in each sample: grade-11 males: CFI: .919, RMSEA: .075 (.073 - .077), grade-12 males: CFI: .914, RMSEA: .075 (.074 - .076), grade-11 females: CFI: .907, RMSEA: .073 (.071 - .075), grade-12 females: CFI: .908, RMSEA: .073 (.071 - .075).[16]

We labelled two of the factors People and Things in males and females, while the last factor was labelled Trades in males and Clerical in females. The labels ‘People’ and ‘Things’ were inspired by Prediger (1982), who provided evidence of two underlying dimensions in the RIASEC hexagon, one of which was termed ‘People/Things’, contrasting Social interests (People) with Realistic interests (Things).

[16] MacCallum, Browne and Sugawara (1996) characterized .080 RMSEA as “mediocre”.

Table 5.4
CFA solution of interests and abilities (grade 12 males)

Variable	Factor		
	Trades	People	Things
I: Trades	.73		
I: Politics		.76	
I: Science			.70
I: Clerical	.42	.55	-.17
I: Medicine		.49	.10
I: Arts		.51	.22
I: Teaching		.71	
I: Military			.45
I: Sales	.13	.78	
I: Architecture	.15		.62
g	-.72		.66
Spatial		-.47	.45
English	-.26	.19	
Speed		.06	
Math		-.11	.07

I = interest scale score. All freely-estimated factor loadings are shown and significant ($p < .001$).

In comparing the three trait complexes to those of PPIK theory, the factor labelled ‘People’ resembled the Intellectual/Cultural trait complex and the ‘Things’ factor resembled the Science/Math trait complex. The People factor had loadings from Artistic interests and the residual English (Verbal) ability as in Ackerman (1996). However, in males particularly there was a greater emphasis on interest scales reflecting Social or Enterprising interests in RIASEC terms (loadings for Politics, Teaching and Sales) than would be predicted for the Intellectual/Cultural trait complex. Therefore, this factor appeared to reflect a broader orientation towards occupations involving interaction with other people (and was thus labelled the People factor). The Things factor had loadings from Science interest and Spatial ability that were consistent with the Science/Math trait complex (Ackerman, 1996). In the males, Spatial ability had the highest positive loading of any residual ability. However, this factor also appeared to be somewhat broader in scope than the Science/Math trait complex, particularly in females where Trades interest also loaded moderately on this factor, consistent with an orientation toward jobs involving manipulation of the physical world (aligning with the Things pole of Prediger's dimension). The moderate loading of Medicine interest on the Things factor in females may seem to contradict this interpretation. However, Science interest had a loading of near-unity on the factor, and the highest correlation of Science interest was with Medicine interest ($r = .59$); thus, the loading for Medicine may have been at least partly an indirect effect of Science interest. This factor also received lower loadings from Math ability than in PPIK theory (Ackerman, 1996).

Table 5.5
CFA solution of interests and abilities (grade 12 females)

Variable	Factor		
	Clerical	People	Things
I: Trades	.27		.50
I: Politics		.73	
I: Science			.99
I: Clerical	.83		
I: Medicine			.61
I: Arts		.72	
I: Teaching		.64	
I: Military		.41	.19
I: Sales	.46	.67	
I: Architecture			
g	-.31	.37	
Spatial		-.18	.20
English		.10	-.14
Speed	.09		
Math			.14

I = interest scale score. All freely-estimated factor loadings are shown and significant ($p < .001$).

The final factor did not resemble the trait complexes proposed by Ackerman and colleagues, but instead seemed to relate to occupational prestige or level of general intelligence. The loadings of *g* on these factors were moderately negative in females, and strongly negative in males. The male factor also obtained a negative loading from English ability. The factor was named 'Trades' in males and 'Clerical' in females due to these being the highest–loading interest scales. Trades and Clerical were occupational interest categories requiring less skilled work and which had lower prestige than the other categories. Although not shown, the structure of the trait complexes was highly similar in the grade-11 samples. The only notable difference was that in the grade-11 samples the factor loadings were marginally lower and the latent classes were slightly less distinct. However, all factor loadings were close in magnitude to those in the grade-12 samples. In males, the loading with the largest difference from grade 12 was for *g* on the Trades factor, which was .14 more positive (loading = -.58 in grade 11). In females, the largest difference was for the loading of Sales on the Clerical factor, which was .05 less positive (loading = .41 in grade 11).

5.3.2 Latent class trait complexes

Latent class analysis was applied to the same interest and cognitive ability scores. The number of classes was decided by examining the changes (decreases) in Akaike information criterion (AIC; Akaike, 1983) and Bayesian information criterion (BIC; Adrian E Raftery, 1995) of the models as additional classes were added, as well as by considering the classification quality metric of entropy (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993).

Examination of the AIC and BIC values showed that they exhibited an "elbow", or levelling off, at five classes in the male samples, and at six classes in the female samples. At this number of classes entropy values were also remained acceptable (close to .80), and the probabilities for most likely class membership were .79 or greater for every class. Thus, We decided to retain five classes in males and six in females. Entropy values were as follows in each sample: grade-11 males (0.762), grade-12 males (0.761), grade-11 females (0.773), grade-12 females (0.767).

Tables 5.6 and 5.7 display the mean standardized values for the interests and cognitive abilities in each latent class in grade-12 males and females, respectively. The latent class means were highly similar in the grade-11 samples, and furnished the same interpretations of the classes; thus they are not shown. The classes varied widely in their mean scores, but the most notable pattern was that two classes contained either people with low occupational interests on all scales (Class 1), or people with high interest on all scales (Class 5). Moreover, mean *g* scores were below average in the low-interest class, whereas *g* level was above the mean in the high-interest class (females) or at mean level (males). This finding was a recurrence of the positive correlation of most of the interest scales with *g*, and the factor-analytic trait complex that related *g* to occupational level or prestige. The remaining classes resembled the factor-analytic trait complexes of People (Class 3 in males, Class 6 in females) and Things (Class 4 in males and Class 3 in females), but this distinction was generally less clear than in the factor-analytic trait complexes. The male results also provided a clearer separation of these two classes than the female samples. As in the factor-analytic trait complexes, the Spatial and English residual abilities sometimes showed an opposing pattern; for example, in grade-12 males Spatial ability was above the mean for the Things-oriented class, but below the mean for the People-oriented class. The Science/Math and Intellectual/Cultural trait complexes could be identified with the Things and People-oriented classes, but as in the factor analysis their interest associations were broader than would be anticipated based on PPIK theory (Ackerman & Heggestad, 1997).

Table 5.6
Latent class means from LCA (grade 12 males)

Variable	Latent class				
	1	2	3	4	5
I: Trades	-.78		-.42	.17	.43
I: Politics	-1.73	-1.03	.41	.31	1.46
I: Science	-1.62	-.18	-.91	.61	.76
I: Clerical	-1.60	-.72	.47	.17	1.15
I: Medicine	-1.32	-.49	-.21	.37	.89
I: Arts	-1.87	-.80	-.23	.42	1.54
I: Teaching	-1.55	-.91	.42	.18	1.46
I: Military	-1.46	-.24	-.34	.36	.74
I: Sales	-2.84	-1.28	.73	.40	1.97
I: Architecture	-2.02	-.39	-.73	.61	1.11
g	-.45			.24	
Spatial		.31	-.90	.29	-.26
English		-.17	.46		
Speed					
Math			-.20		

Note: Class means between -.14 and .14 not shown. Composition of sample: class 1 = 8.4%, class 2 = 26.0%, class 3 = 15.1%, class 4 = 34.4%, class 5 = 16.0%.

Table 5.7
Latent class means from LCA (grade 12 females)

Variable	Latent class					
	1	2	3	4	5	6
I: Trades	-1.04	-.35	2.33	-.46	2.32	
I: Politics	-1.38	-.38		-.50	2.02	1.69
I: Science	-1.27	-.60	1.06	.72	1.89	.80
I: Clerical	-.46	.60	.27	-1.12	.90	
I: Medicine	-.92	-.47	.32	.86	.96	.58
I: Arts	-1.33	-.16	.38		1.25	1.02
I: Teaching	-1.07				1.25	1.02
I: Military	-.95	-.22	.77	-.20	1.47	.55
I: Sales	-1.74	.43	.38	-.99	2.05	1.19
g	-.19		.33	.92	.42	.82
Spatial			.40	.19		
English			-.20		-.20	
Speed						
Math				.26	.15	.20

Note: Class means between -.14 and .14 not shown. Composition of sample: class 1 = 21.4%, class 2 = 25.8%, class 3 = 12.0%, class 4 = 13.3%, class 5 = 7.6%, class 6 = 19.8%.

5.3.3 Prediction of occupational type

In the last stage of the analysis, the cognitive ability and interest scores were used to predict the occupational type of participants eleven years after high school. Tables 5.8 and 5.9 display the grade-12 results for the odds ratios for the logistic regressions of the individual scores and factor-analytic trait complexes (entered in two separate analyses). All predictors were standardized, hence the odds ratios represent the increase/decrease in the odds of attaining the particular occupation type given a one standard deviation increase in the variable. Due to the large sample sizes, confidence intervals for odds ratios were very small and are not shown.[17] To examine the predictive validity of the individual scores and trait complexes, the sizes of the odds ratios and pseudo R^2 values were compared. The results for the grade-11 samples were similar and not shown.

[17] For example, in the prediction of science jobs in grade-12 males, the mean 95% confidence interval occurred from .022 below the odds ratio estimate (SD = .017) to.023 above the estimate (SD= .017).

The odds ratios for the individual scales were consistent with the previous research on the predictive validity of cognitive abilities and interests (Schmidt & Hunter, 2004; Wai et al., 2009). According to the odds ratios, the strongest cognitive predictor for most categories was general intelligence, but the residual abilities also made some notable contributions. For example, the English residual ability had strong associations with the Fine Arts categories in both males and females. The residual abilities also showed discriminative predictive validity, where higher Spatial ability, for example, contributed positively to scientific and technical jobs, but led to lower probabilities of attaining social-oriented jobs (such as Teaching, and Humanities in males). The interest composites also showed good discrimination among categories; for example, interest in Teaching was strongly predictive of attaining a job in that category, but negatively related to attaining jobs in several other categories.

The odds ratios for the trait complexes were generally consistent with their effectiveness in capturing the shared variance between the interests and cognitive abilities. The Things factor had large effects on the probabilities of entering in scientific and technical jobs, whereas the People factor affected jobs in social-oriented categories. Greater scores on the Trades and Clerical factors decreased the odds of being in the professional job categories, and increased the odds of being in the non-professional and semi-skilled job categories (such as Construction jobs in males, and Clerical jobs in females).

Table 5.8

Odds ratios of abilities and interests predicting job categories (grade 12 males).

Predictor	Job Category										
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
g	3.71	3.78	1.25	1.49	2.66			.81	.72		.61	.48
Spatial	1.32			.81	.69	1.34				.83	1.23	
English	.76				1.20	1.67						
Speed				1.45		.75				1.46		
Math	1.39	1.63			.70	.82			.83			
I: Trades			.65		.66	.64	.86	.82	1.55		1.88	1.51
I: Politics	1.21				1.30	1.17			.82			
I: Science	2.41	.59		.72	.84		1.49					
I: Clerical			1.19		.63	1.36	1.42			1.79		
I: Medicine		2.92				.85				.68		.82
I: Arts		.76	.83		1.17	1.80	1.22	.68	.82		1.36	1.22
I: Teaching	.84	.79		2.66	1.19	.52	.82	.82	.80	.76	.72	
I: Military								1.19			.79	
I: Sales	.74		1.28		1.45	.83	.61	1.38	.82			
I: Arch.									1.21	.84		.80
F1: Trades	.27	.19	.58	.80	.20	.47	.70		2.52	1.66	3.46	3.87
F2: People	.44		1.37	2.33	2.28		.59		.43	1.20	.66	
F3: Things	5.72	2.97	1.40	.67	1.48	1.60	1.89			.54	.65	.29
R^2: full	.52	.47	.11	.25	.42	.21	.10	.06	.22	.15	.26	.25
R^2: factors	.37	.39	.11	.15	.46	.11	.06	.00	.29	.06	.29	.32

Note: Odds ratios between .86 and 1.14 not shown. R^2: full was for individual scales as predictors; R^2: factors was for the trait-complex factors. Med. = Medicine, Mech. = Mechanical.

Table 5.9

Odds ratios of abilities and interests predicting job categories (grade 12 females).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
g	3.85	1.94		1.74	2.97	.64	1.59	.70	.58		.41	.50
Spatial		1.17		.82		1.18			1.16			
English	1.23				1.17	1.49	1.35	1.26		1.17	1.35	
Speed			1.27		1.35	.78		1.19				
Math	2.13		1.25	1.28	1.46	.67	.75			.85	.24	.77
I: Trades	.61	.85		1.18		.27			1.32		.39	1.23
I: Politics	1.30	1.28			1.24	1.70	.85	.75	.75		.43	
I: Science	3.92		1.31	1.20		.78	1.73		1.58			
I: Clerical				.66	.47	.56	2.17		.75	1.35	.47	
I: Medicine	.58	3.22	.64			1.32		.79	.63			
I: Arts	.55	.84	1.13		1.50	3.31		1.26	1.19		.74	
I: Teaching	.85	.79	.64	1.89						.82	.73	
I: Military	1.13				.76	1.45		1.19	.69		1.52	
I: Sales	.81	.85	1.64		1.72		.75	1.35	.90	1.20	2.07	
F1: Clerical	.43	.46	1.52	.40	.28	.28	1.57	1.42		1.61	.56	1.33
F2: People	.65			3.25	6.62	7.21	.60		.39		.70	.40
F3: Things	5.71	2.84	1.18		.46	.35	2.16	.65	1.77		.53	1.61
R^2: full	.55	.39	.13	.29	.45	.52	.21	.13	.19	.08	.63	.18
R^2: factors	.45	.30	.05	.29	.42	.42	.11	.07	.13	.06	.30	.11

Note: Odds ratios between .86 and 1.14 not shown. R^2: full was for individual scales as predictors; R^2: factors was for the trait-complex factors. Med. = Medicine, Mech. = Mechanical

The variances explained by the individual scale scores and the trait complexes were equal or nearly equal for some categories (such as Humanities in males and females, and Teaching in females), but were generally lower for the trait complexes. For the grade-12 males, the mean pseudo-R^2 for the trait complexes was 21.8% (SD = .15), compared with 25.2% (SD = .15) for the full scores. In females it was 22.6% (SD = .15) compared with 31.2% (SD = .19) for the scale scores. Therefore, our hypothesis that the trait complexes would show predictive validity equal to the individual scores was not supported. The same conclusion was drawn for the grade-11 results (see Supplemental Tables C1 and C2).

Tables 5.10 and 5.11 display the odds ratios from logistic regressions of the jobs categories onto latent class memberships in grade-12 males and females. The reference class was chosen as the largest group (class 4 in males and class 2 in females). The odds ratios in the male data were generally smaller than for the individual scores or factor-analytic trait complexes. However, in grade-12 females, several large odds ratios were observed for the probabilities of attaining Science, Medicine and Fine Arts jobs. This was likely due to the small frequencies of jobs in these categories for females, such that those who attained them were outliers in interests and abilities. The mean pseudo-R^2 for the LCA trait complexes was 6.3% (SD = .04) for grade 12-males, and 13.7% (SD = .09) for the grade-12 females. These values were considerably lower than the variance explained by the scale scores (the same pattern occurred in the grade-11 samples; see Supplemental Tables C3 and C4). The higher explained variance in females was likely partially attributable to the use of one additional class. The hypothesis of equal predictive validity was clearly rejected for the latent-class trait complexes.

Table 5.10
Odds ratios of latent classes in predicting job category (grade 12 males).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.[a]	Labour
Class 1	.33	.16	.37	.64	.39	1.89		.76	1.80	.31	1.68	2.96
Class 2	.69	.41	.47	.74	.80	1.33			1.26		1.54	2.05
Class 3	.10	.57		3.60	2.30	.50	.28		.45	1.17	.51	1.37
Class 5	.63	.40		2.10	1.97	1.76			.67	.68		
R^2	.15	.08	.05	.10	.07	.05	.06	.00	.05	.04	.04	.04

Note: Reference class is class 4. Odds ratios between .86 and 1.14 not shown. Med. = Medicine, Mech. = Mechanical

Table 5.11
Odds ratios of latent classes in predicting job category (grade 12 females).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.[a]	Labour
Class 1	3.90	1.57	.52	.78	.34	3.36	.44		3.24	.73	n/a	1.99
Class 3	9.56	2.47	2.57	1.41	1.39	1.60	.71	1.25	1.41	.66	n/a	2.27
Class 4	15.60	15.11	.35	2.40	1.58	3.38	.29	.12	2.91	.46	n/a	
Class 5	11.89	6.26	.77	2.82		1.36		.32	1.34		n/a	
Class 6	18.05	4.90	1.16	2.94	2.83	25.52		.39		.61	n/a	.24
R^2	.29	.20	.09	.08	.13	.28	.06	.14	.08	.02	n/a	.14

Note: Reference class is class 2. Odds ratios between .86 and 1.14 not shown. Med. = Medicine, Mech. = Mechanical
[a] model did not converge.

5.3.4 Multinomial prediction

An additional way to predict the occupational categories in Project TALENT was using multinomial logistic regression, where the binary outcomes of the twelve categories were predicted simultaneously. An advantage of this method was that the ability to classify individuals into the correct occupation category could be determined, comparing across the three different sets of predictors (the individual scores, the factor-analytic trait complexes, and the latent-class trait complexes). Austin and Hanisch (1990) similarly examined the classification accuracies of their five discriminant functions in the grade-10 sample of Project TALENT; therefore our results can also be compared to theirs.

Tables 5.12 and 5.13 display the correct classification percentages from the multinomial regressions for the grade-12 males and females (the analysis was not performed for the grade-11 samples). The percentages are provided for the three different sets of predictors. The correct classification percentage was greater than chance when it exceeded the sample percentage (the percentage in the population who were in that occupation category); hence these values are given for comparison. The sample percentages differ slightly from those in Table 5.1 because the “Vague and Undesigned” occupation category was excluded (since it would not be expected to be able to classify individuals in that category).

The mean classification accuracies are given in the bottom rows. The mean accuracies indicated that the individual scores had the highest classification accuracy, followed by the factors and then the latent classes. The classification accuracy exceeded chance with the individual scores for nine of the twelve categories in males and six of twelve in females. This accuracy was reduced when the factors or classes were used as predictors. This finding replicated the results from the logistic regression that the factors and classes had lower predictive power than the individual scores for cognitive abilities and interests.

Table 5.12
Original sample composition and correct classification percentages for multinomial regression (grade 12 males)

Job category	Sample percentage	Classification success (% correct)		
		Individual Scores	Factors	Classes
Science	6.2	42.0	17.8	0
Medicine	2.8	16.4	0	0
Business	21.3	62.0	74.2	79.0
Teaching	9.1	24.4	1.9	0
Humanities	3.9	16.5	3.0	0
Fine Arts	1.0	0	0	0
Technical	5.7	0.4	0	0
Sales	11.9	9.1	0	0
Mechanical	9.6	20.0	15.5	0
Clerical	8.5	19.8	1.4	0
Construction	3.8	3.0	0	0
Labour	16.2	61.0	61.6	49.1
Mean	8.33	22.9	14.6	10.6

Table 5.13
Original sample composition and correct classification percentages for multinomial regression (grade 12 females)

Job category	Sample percentage	Classification success (% correct)		
		Individual Scores	Factors	Classes
Science	0.4	17.6	0	0
Medicine	7.3	35.3	3.5	0
Business	7.3	0.3	0	0
Teaching	19.2	55.8	42.4	20.2
Humanities	2.1	1.7	0	0
Fine Arts	1.0	27.8	0	0
Technical	4.5	0	0	0
Sales	4.5	0	0	0
Mechanical	1.1	0	0	0
Construction	0.04	0	0	0
Clerical	34.6	74.7	80.0	92.0
Labour	18.0	38.7	17.2	0
Mean	8.33	21.0	11.9	9.4

5.4 Discussion

Cognitive abilities and occupational interests are intertwined, and several developmental theories have been advanced to explain these associations (Ackerman, 1996; Gottfredson, 1986; Hogan & Roberts, 2000). Only PPIK theory, however, has provided hypotheses of the overlap between particular interests and abilities, which are said to be captured by two trait complexes termed Math/Science and

Intellectual/Cultural (Ackerman, 1996; Ackerman & Heggestad, 1997). Trait complexes have been proposed as important influences on occupational knowledge, and thereby career choice (Ackerman, 1996; Ackerman & Beier, 2003a). Yet, previous research had not examined whether trait complexes can predict future occupation. In addition, Ackerman and colleagues have relied on factor analysis to extract trait complexes, though latent class analysis is arguably more consistent with their definition. We reasoned that if PPIK theory was correct then trait complexes obtained from interest and cognitive ability scores in Project TALENT would fulfil two conditions. First, the content of the Science/Math and Intellectual/Cultural trait complexes would be replicated, and second, they would show equal or greater predictive validity than individual scales score for predicting occupational type. The first condition received only mixed support and the second was not supported; we discuss each of the hypotheses in turn.

When an acceptable confirmatory factor analytic model of the trait complexes was constructed, factors were found that resembled the Science/Math and Intellectual/Cultural trait complexes. However, the involvement of interests in the trait complexes was broader than proposed in PPIK theory, and the factor content aligned more closely with the two poles of the People/Things interest dimension (Prediger, 1982). The People factor had loadings from Sales, Politics and Teaching interests, which correspond to Enterprising and Social interests in the RIASEC framework. These interests were not hypothesized to be part of the Intellectual/Cultural trait complex (Ackerman & Heggestad, 1997), but are characteristic of the People interest pole (Prediger, 1982). The Things factor displayed a closer correspondence with the Science/Math trait complex, except that the loading of residual Math ability was minimal. Johnson and Bouchard (2009) also found that a broad section of people-oriented interest groups had higher verbal abilities, while groups that were things-oriented displayed higher spatial (image rotation) abilities. This also contradicted the more narrow focus of PPIK theory. However, in this study, the most notable departure from PPIK theory was the presence of a third factor.

The factor captured the negative associations of general intelligence with the Trades and Clerical interests. Tracey and Rounds (1996) found a third dimension of the RIASEC interests that is relevant to this finding. The first two dimensions were defined by Prediger (1982), who labelled them People/Things and Data/Ideas (Data/Ideas was oriented between Conventional and Enterprising interest (Data) and Investigative and Artistic interests (Ideas). Tracey and Rounds (1996) performed a principal component analysis of interest ratings for 229 occupation titles, and found a third component that was related to occupational prestige or socioeconomic status. Our third factor was consistent with such a prestige dimension, and specifically included *g,* which Tracey and Rounds did not measure. The involvement of *g* in the trait complexes was extensive. Considered together, the trait complexes explained by far the most variance in the *g* factor. For example, in the grade-12 males, the trait complexes explained 43.7% of the variance in *g*, but only 18.9% for the next-highest ability (Spatial ability). In fact, there was more variance explained in *g* than in the four specific abilities combined, which had a total of 27.6% of their variance explained by the trait complexes. In the grade-12 females, this distinction was even stronger, with 21.9% of the *g* variance explained, and 7.2% for the specific abilities combined. These ratios were similar in the grade-11 samples. In the latent class analysis, it was observed that classes which reflected greater interests in higher-status occupations (such as Science or Politics) also had above-average *g* levels. General intelligence was notably involved with two of the five LCA trait complexes in males, and five of six in females, while the specific abilities generally played lesser roles.

In the PT scales, there was a moderate general interest factor, and it was associated with *g*, particularly according to the LCA results. Some researchers have cautioned that the general interest factor is likely a “response set” owing to acquiescence bias or other methodological factors, and have advised to control for it (e.g. Prediger, 1982). Some occupational interest scales intrinsically control for differences in average level of response by ipsatisation. However, other researchers have suggested that the general interest level could have some substantive psychological meaning, noting that it correlates positively with Extraversion and Emotional Stability (see Rounds & Tracey, 1993, and the references therein).

The Science/Math and Intellectual/Cultural trait complexes of PPIK theory emphasize the importance of specific abilities in interest-ability associations, but we found that *g* played a more important role in the trait complexes than did specific abilities. These findings replicated those of Johnson and Bouchard (2009) that there were substantial differences in mean *g* level across latent-class interest groups, in line with their average occupational status. The results lend support to Gottfredson's theory of circumscription and compromise, which specifies that *g* plays a central role in determining the occupations in which individuals become interested, according to the levels of education and training required (Gottfredson, 1986, 2005).

PPIK theory specifies the involvement of intelligence-as-process in the Science/Math trait complex (where intelligence-as-process could be considered similar to *g*), but does not propose direct involvement of consideration of social status or training requirements in the emergence of the trait complexes. One possible reason for this is that Ackerman and colleagues have used RIASEC measures of interests, which are limited in their representation of low-prestige occupations (Deng et al., 2007; Tracey & Rounds, 1996). In contrast, the Trades and Clerical interest scales derived from PT items were primarily formed from low-prestige occupation titles. Moreover, as noted in the introduction, Ackerman and colleagues did not separate *g* variance from specific-ability variance in measuring cognitive abilities, and thus were not able to evaluate the roles of *g* and specific abilities separately in their trait complexes. In addition, they have often used samples of college students, which suffer from range restriction of cognitive ability, as well as occupational interests (Ackerman, 2000; Kanfer, Wolf, Kantrowitz & Ackerman, 2010).

The People and Things factors were consistent with the People/Things interest dimension. However, the factors were positively correlated, which suggests that they did not act as poles of one dimension in the current study. This may be attributable to differences in the interest scales used. Studies that assess the RIASEC types typically find that the correlations among types vary from moderately positive to moderately negative (De Fruyt & Mervielde, 1999). In the PT data, all the interest composites correlated positively. This shared variance between interest scales made it unlikely to obtain factor-analytic trait complexes would that were negatively correlated with each other, given that the interest scales made up the majority of the variables entered in the analysis.

One source of the more positive correlations among the PT scales compared with the RIASEC measures could have been that the RIASEC scales were designed to emphasize the separation between the types by selecting occupation titles that are unambiguous representatives, whereas the PT occupation titles were not pre-selected to fit separate categories. Thus, a substantial portion of each item response was made up of the student's general level of occupation interest. Another methodological factor that could have contributed to this common variance was acquiescence due to testing fatigue because the participants were young and were required to complete many scales during the course of the study (Flanagan et al., 1962). Youth could have contributed substantively to the general interest factor as well, given that the students may not have been aware of the challenges in different occupations and thus responded more positively to a wide variety of titles than older and more knowledgeable respondents would have done. Thus, the data in the current study were probably not ideal to assess whether People/Things consists of one bi-polar dimension or two dimensions, although the results did suggest two separate dimensions. Overall, the People and Things factors that we observed were generally consistent with Intellectual/Cultural and Science/Math trait complexes, but not identical to them. While the cognitive abilities generally showed the expected relations with the two factors, the interest loadings appeared to capture divisions that were more consistent with Prediger's People/Things distinction than the distinction between Cultural and Scientific interests emphasized in PPIK theory.

The structures of the trait complexes using both methods were very consistent across grades. The structure was slightly less clear structure in grade 11, but this could have simply resulted from the interests and cognitive abilities being less developed and differentiated in the younger sample. The trait complexes were less consistent between genders than grades, but in both groups three factors were obtained that were recognizable as People, Things and prestige factors. In the latent class analysis, the numbers of classes obtained for males and females differed by one, but both sets of classes captured groups that were organized according to occupational level and the People-Things distinction. The prestige factor for females primarily related *g* negatively to Clerical interests, while in males it related *g* negatively to Trades interests. This difference is consistent with the finding that males are on average more interested in Realistic-type occupations which Trades fall under, while females are more interested in Social-type interests, which are relevant to the Clerical factor (Deng et al., 2007). However, social roles for men and women were also likely involved in this difference in which lower-prestige occupations they preferred. In the 1970's it was very uncommon for women to enter Realistic-type occupations, and less likely for men to enter Social-type occupations that were Clerical in nature, compared to other non-Clerical jobs. The limitation of the time period in regard to sex differences is addressed further below. In summary, the trait complexes obtained were consistent across samples yet differed from those predicted in PPIK theory. The present study went beyond previous research in studying the associations of interests to cognitive abilities, to the prediction of attained occupation. The individual scores for cognitive abilities and interests had substantial power in predicting occupational category eleven years after high school, consistent with previous studies (J. T. Austin & Hanisch, 1990; Humphreys et al., 1993). However, this was the first study using PT data to find that specific abilities, independent of g, also predicted some occupations. Most notably, residual Spatial and English abilities displayed the clearest patterns of prediction, with Spatial ability predicting scientific and technical jobs and English ability predicting jobs in the broad Humanities area. Residual Math and Speed abilities were less consistent predictors, but Math ability, for example, was predictive of future Science jobs in both males and females.

Trait complexes derived by factor analysis were strong predictors of occupational type, but, excepting a few occupational categories, they explained less variance than individual scores for cognitive abilities and interests. This was the case using both logistic and multinomial regression. The latent-class trait complexes performed notably worse than the factor-analytic trait complexes in predicting occupational type.

The weaker predictive ability of latent-class trait complexes suggests that trait complexes, if they exist, should not be conceptualized as groups of individuals who share similar levels of interest and cognitive ability variables. Rather, there was greater support for the idea that trait complexes could be conceptualized as capturing parts of the shared variance between cognitive abilities and interests, where these variables and the trait complexes are continuous. This shared variance may be the result of the reciprocal influences of cognitive abilities and interests upon each other through development, as theorized by a number of researchers (Ackerman, 1996; Armstrong et al., 2008; Hogan & Roberts, 2000), although direct evidence for this is still lacking.

One possible limitation of the current study was that the trait complexes that primarily involve personality-intelligence associations were excluded (the Social and Clerical/Conventional trait complexes). Nonetheless, Ackerman and Beier (2003) put forth the general claim that trait complexes are more informative about career choices than individual scales, which should apply to all their trait complexes. In addition, previous research on trait complexes has found the Science/Math and Intellectual/Cultural trait complexes to be the two most important predictors of specialized knowledge (Ackerman & Rolfus, 1999) and the university course that students select (Ackerman, 2000). Nonetheless, future research could be done to address the predictive validity of the Social and Clerical/Conventional trait complexes for attained occupation.

A second limitation was that the selection of the number of trait complexes through EFA and LCA was subjective, and guided in large part by their interpretability. The confirmatory models also displayed marginal fit, although the use of CFA is an advance over previous studies of trait complexes that have only used exploratory methods. The difficulty in constructing factor models for the trait complexes may have been exacerbated by problems with the interest and cognitive ability measures. As discussed above, the interest scales displayed a positive manifold, which is inconsistent with research on the RIASEC that has found a circumplex structure for interests (Armstrong et al., 2008; Holland, 1997). The amount of variance for which the residual abilities accounted was small, which was at least partly attributable to the modest number of tests for each. This may have contributed to the weaker involvement of the residual abilities in the trait complexes compared to *g*, as well as their lesser predictive validity. However, previous studies have also found that g has the greatest importance (Gottfredson, 1986; Johnson & Bouchard, 2009).

Another limitation of research on future occupation is that the power of prediction is dependent on the job market. If the skills and interests in the population do not match the requirements for the available jobs, at least some individuals will be mismatched. In addition, there are social and economic pressures that may act to lead individuals away from their ideal occupations. The 1970s was a period of increasing educational opportunity in the United States, but the occupational opportunities were not as dominated by educational qualifications as in the present day. For women, strong social expectations about gender roles in the division of labour were present: nearly half the women at follow-up were housewives, and the most prevalent paid occupation was in the Clerical and Office Work category. Gender expectations prevented many women from selecting jobs for which their cognitive abilities and interests were suited. However, the total predictive validities of abilities and interests were not substantially lower for women compared with men, even for male-dominated occupation categories. This was possibly because men with a wider range of abilities and interests would have obtained these jobs, thus diluting the predictive validity of the baseline variables. One notable divergence was that the explained variance was greater for Fine Arts occupations for females than males, which suggests that gender roles may have also restricted the occupational opportunities of men. In comparison with the present day, manufacturing and trades jobs were more prevalent, which provided a niche for more low-*g* workers; the predictive validity of *g* and trades' interests may be lower in modern samples relative to PT.

The mean classification accuracy across the twelve categories in Austin and Hanisch (1990) was 30.5%. Here it was lower, even with the use of full scores (22.9% in males, 21.0% in females). There were several factors that likely contributed to this. First, Austin and Hanisch did not separate their sample by gender, and instead entered it as a variable in the discriminant function analysis. As the genders differed strongly in their frequencies across occupation category (see Table 5.1), this would have increased their predictive power compared to ours. Socioeconomic status was also used as a predictor in that study but not in the current one.

In addition, a limitation of multinomial regression is that unequal proportions in the outcome variable decreases the prediction accuracy because individuals are more likely to be classified into the more common prior categories. For example, the two most common categories in males, Business and Labour, were overrepresented in the classifications. The grade-10 sample of Project TALENT was more evenly distributed amongst the twelve categories than the grade-11 or grade-12 samples, which likely contributed to the greater classification accuracies found by Austin and Hanisch (1990).

In summary, our first finding was that the Science/Math and Intellectual/Cultural complexes in PPIK theory could not be closely replicated because the trait complexes we found were broader in content and gave much more weight to *g*. Within the factor-analytic trait complexes, *g* had the most explained variance, and one factor related low *g* to interest in Clerical and Trades occupations (identified with the prestige dimension of Tracey and Rounds [1996]). These findings were consistent with the theory of Gottfredsson (1986, 2005) that *g* acts as an important filter in occupations according to status level. Neither type of trait complexes were equal predictors of attained occupation when compared to individual traits, which calls their theoretical status into question. The greater predictive validity for factor-analytic trait complexes than latent-class trait complexes suggests that the Science/Math and Intellectual/Cultural trait complexes, if their definition is expanded, may be useful summaries of the overlap between cognitive abilities and interests, but they do not appear to represent discrete groups in the population with combinations of different trait levels.

Chapter 6: Conclusion

The studies in this thesis examined the links among cognitive ability, personality and occupational interests in Project TALENT. The research built towards testing of a key component of the integrative framework of Ackerman and colleagues (Ackerman, 1996; Ackerman & Beier, 2003a; Ackerman & Heggestad, 1997), the concept of trait complexes made up of cognitive abilities and interests. In this final chapter, the main results of the studies are summarized, and their implications are discussed within the context of research that strives to integrate the three domains of individual differences. Some of the limitations of PT data to address this topic are discussed, and suggestions for future research are provided.

5.1. Cognitive ability

The study presented in chapter 3 was designed to investigate the psychometric structure of the cognitive ability tests in PT. Three of the most well-regarded models were compared: the VPR model, the CHC and the Extended Gf-Gc models. The VPR model was found to have the best fit to the test data in all samples. The results provided replication of three previous model comparison studies where the VPR model outperformed the CHC and Gf-Gc models (Johnson & Bouchard, 2005a, 2005b; Johnson, Te Nijenhuis, et al., 2007). This comparative research has suggested that the VPR model is the most accurate representation of human cognitive abilities, thus it follows that this model should be the best suited for understanding the overlap among cognitive abilities, personality and interests. Nonetheless, the main purpose of this study in the thesis was to develop a model appropriate for use in summarizing the cognitive ability measures in PT for further research. The topic of how the VPR model can contribute to integrative research is discussed in section 5.3 below.

5.2. Personality-intelligence associations

The study presented in chapter 4 investigated linear and nonlinear associations between general intelligence and personality. The linear associations that were observed were in line with previous research: *g* was positively associated with PT scales reflecting Openness to Experience and negatively related to Neuroticism scales (Ackerman & Heggestad, 1997; Zeidner and Shani-Zinovich, 2011). There was mixed support for the hypothesis of a negative association between *g* and Conscientiousness: *g* was negatively associated with Tidiness, but not Maturity (Chamorro-Premuzic & Furnham, 2006). Finally, the results supported Ackerman and Wolf's (2005) hypothesis that the social potency aspect of Extraversion is positively associated with *g*, while social closeness is negatively associated.

In contrast to most previous studies of nonlinear associations, several significant quadratic effects of *g* on personality traits were also found. These quadratic associations were predicted primarily on the basis of previous research with gifted (high *g*) samples (Sak, 2004; Zeidner & Shani-Zinovich, 2011). Three of the most consistent nonlinear associations were between greater *g* and greater social potency (Leadership), lower social closeness (Sociability), and lower scores on the Tidiness facet of Conscientiousness. Another conclusion drawn from the study was that the general factor in personality self-ratings may be an important confound to consider when studying personality-intelligence associations.

These findings can be applied to integrative research. PPIK theory and the integrative framework of Armstrong and colleagues have only addressed linear associations between cognitive abilities and personality (Ackerman & Heggestad, 1997; Anthoney & Armstrong, 2010; Armstrong et al., 2008). This limitation is due to the methods used to identify associations between the three domains: factor analysis and multidimensional scaling, which only capture linear associations among variables. However, even within linear associations, neither theory has taken into account the negative associations of *g* with Neuroticism and the Tidiness aspect of Conscientiousness, as well as the differential link of *g* with the two aspects of Extraversion. These links, although smaller than the relation between *g* and Openness to Experience, could be important in understanding how *g* transacts with personality to influence the development of occupational interests. For example, individuals with higher *g* scores may be less interested in occupations in the Conventional RIASEC domain, not just because of the occupations' lower average prestige, but because Conventional jobs have a high requirement for Conscientiousness, which includes Tidiness (Armstrong et al., 2008).

Although the methods used by existing theorists have not been suited to incorporating nonlinear associations, one possible method to do so is to examine groups selected by extreme g scores. The results from chapter 4 suggest that a group defined by high *g* would have outlying scores on the personality traits that demonstrated nonlinear discontinuity with higher *g* (higher social potency, lower social closeness and lower Conscientiousness). Understanding the set of characteristics that are specific to groups with high *g* or low *g* levels could also be important for an integrative theory. For example, the higher average social potency of the intellectually gifted may incline them towards Enterprising occupations, but their lower average social closeness may lead them away from Social occupations that involve working personally with others. Such influences may be underestimated if only linear associations between *g* and personality are considered. This conclusion only addresses the associations between *g* and personality traits, but nonlinear associations may also exist between domain-specific cognitive abilities and personality traits. In fact, nonlinear associations may exist among all three of the domains, but this is beyond the scope of the current analyses.

Issues surrounding the “general factor of personality” are also a concern for integrative research. Across the eight grade and gender samples in PT, this personality factor displayed a mean correlation of .28 with *g*, an association which has also been observed in several other studies (Dunkel, 2013; Irwing, Booth, Nyborg, & Rushton, 2012; Loehlin, 2011). In such cases, the “lower-order” personality traits will all tend to be positively correlated with *g*, though the associations may be non-significant or negative when the GFP is controlled. Future research is needed to determine if the association between *g* and the GFP involves genuine personality variance or some artifactual source, as the standing of the GFP is still in question (Chang et al., 2012; Hopwood, Wright, & Brent Donnellan, 2011). In light of this issue, studies that examine the associations of personality and cognitive abilities should control for socially-desirable responding, ideally by using multiple raters instead of social desirability scales, as social desirability scales also contain substantive variance (Paulhus, 2002). The presence of a strong GFP in the PT personality scales, with no way to control for socially desirable responding, is a significant limitation of this dataset for integrative research.

5.3. Trait complexes

In chapter 5, the concept of trait complexes of interests and cognitive abilities was tested by examining their long-term predictive validity. It was found that three trait complexes obtained by factor analysis had nearly the same predictive power for occupation as individual scores of interests and cognitive abilities, while trait complexes obtained by latent class analysis performed substantially more poorly than either. As latent class analysis is more consistent with the definition of trait complexes by Ackerman and colleagues (Ackerman & Beier, 2003a; Ackerman & Heggestad, 1997), this definition was undermined. Instead of being viewed as combinations of levels of traits that exist in certain groups of the population, trait complexes could only be defended as reflecting shared variance among continuous variables. This primary conclusion is in line with the view of Armstrong and colleagues that the integration of cognitive abilities, interests and personality is to be best understood by considering the associations among dimensional variables (Anthoney & Armstrong, 2010; Armstrong et al., 2008).

The People and Things factors that were obtained were broader in their content than was predicted from PPIK theory (the Science/Math and Intellectual/Cultural trait complexes). This result was in line with previous findings by Johnson and Bouchard (2009) regarding the distribution of verbal and spatial abilities across interest groups. In both that study and in chapter 5, verbal ability was associated with broadly people-oriented interests, while spatial ability was aligned with things-oriented interests. Moreover, in chapter 5, I found that factors representing these overlaps were differentially predictive of technical-scientific versus artistic-humanities jobs. When considered in the context of the VPR model, these findings strongly suggest that the Verbal-Image Rotation dimension is aligned with the People/Things dimension first proposed by Prediger (1982). The Verbal/Spatial distinction in cognitive abilities may be key to understanding how the People/Things dimension in interests emerges. Differential success in dealing with verbal and spatial tasks may influence individuals' future interest levels in People versus Things occupations because of their different task demands. The People/Things dimension is defined by interest in working closely with other people versus working with physical objects and data. However, People-oriented occupations are more likely to require verbal communication abilities, while Things-oriented occupations are more likely to require skill at spatial manipulation. Thus, through perceptions of their own verbal and spatial abilities, and their knowledge of job activities, individuals are likely to gravitate towards either People or Things-type occupations.

In addition to the important role of the Verbal-Image Rotation dimension in the VPR model, Johnson and Bouchard (2007) found that it is the primary dimension along which sex differences occur in cognitive abilities. This is consistent with its association with the People/Things dimension, as the sex difference on this dimension is one of the largest among psychological traits (Armstrong et al., 2011; Lubinski, 2000). There is some evidence that the sex differences in both interests and cognitive abilities could be driven by genetic differences that are traceable to different evolutionary investment strategies for males and females (see Johnson & Bouchard, 2009, and the references therein). However, in Johnson and Bouchard (2009), sex differences in occupational interests were found to be larger than those for specific and general cognitive abilites, suggesting that socialization pressures may cause a greater separation in interests than would be expected on the basis of cognitive abilities alone (Johnson & Bouchard, 2009).

Along with its better fit over rival models, the central role for sex differences in cognitive ability in the VPR model is an important element that supports its use for integrative research. In contrast, these differences are less well-articulated in theories for the CHC model (Horn & Blankson, 2005) and Gf-Gc models (McGrew, 2009). For example, in the latest review of CHC theory, sex differences were not even mentioned (McGrew, 2009). In these models, the distinction between spatial and verbal abilities is typically confounded with the distinction between fluid and crystallized factors, which undermines the investigation of sex differences. For example, tests of crystallized intelligence that rely on general knowledge have often been found to favour males (Keith, Reynolds, Patel, & Ridley, 2008). However, tests that require specific verbal knowledge, such as spelling and vocabulary, tend to favor females (Johnson, Bouchard, et al., 2007). Therefore, focusing on crystallized intelligence rather than verbal intelligence may lead researchers to overlook a salient sex difference in cognitive ability. PPIK theory still refers to the Gf-Gc model (Ackerman, 1996), and Armstrong and colleagues have not explicitly adopted any intelligence theory (Armstrong et al., 2008). The results of chapter 5 suggest that both theories could benefit from incorporation of the VPR model of cognitive abilities.

Another crucial oversight in PPIK theory is that it does not include the prestige dimension in occupational interests and its association with *g*. In PT, the involvement of *g* in the Science/Math trait complex was substantial, as predicted in PPIK theory, but *g* was also related to greater overall occupational interests, and particularly to interests in occupations with higher prestige. The strong association of *g* with higher-prestige interests supported the developmental theory of Gottfredson (Gottfredson, 1986, 2005).

The three "trait complex" factors found in the study tended to support the RIASEC-based approach of Armstrong and colleagues (Armstrong et al., 2008). As discussed above, two of the factors appeared to align with the People/Things dimension that underlies the RIASEC (Armstrong et al., 2008; Prediger, 1982). The third factor was related to occupational prestige or level, a dimension which has also been incorporated into their framework (Armstrong et al., 2008; Deng et al., 2007). However, typical RIASEC measures such as the Vocational Preference Inventory do not contain enough breadth of occupations to cover the whole prestige dimension (Deng et al., 2007). In addition, Armstrong and colleagues have used aptitude ratings instead of objective intelligence tests to assess cognitive abilities. Further research is needed to replicate the association of this third interest dimension with *g*.

A limitation of the study in chapter 5 was that personality was not included, in order to focus on first integrating cognitive abilities with interests. It is possible that the trait complexes in PPIK theory that involve personality-interest associations could be found in PT, and that they may display better predictive validity than the trait complexes for interests and cognitive abilities. However, as observed in chapter 4, the PT personality measures have greater limitations for this research than do those for the interests, because of the lack of item-level data and the presence of a large common factor.

5.4. Suggestions for future research

The research presented in this thesis has provided several important indications for future integrative research. First, the studies in this thesis support the conclusion that the VPR model of cognitive abilities is likely the best existing model of cognitive abilities for this integrative research. This is due to its better description of the structure of cognitive abilities, and the strong links between the Verbal-Image Rotation dimension and the People/Things dimension of occupational interests. In addition, future studies should separate *g* from specific abilities in their intelligence models (i.e. specify them to be uncorrelated factors), as this is the only way to accurately assess their relative contributions in integrative models.

Second, integrative research could benefit from the examination of nonlinear personality-intelligence associations, which were responsible for some substantive personality differences of individuals at high and low *g* levels. These personality differences, such as lower social closeness and tidiness, and higher social potency, could be relevant to the development of occupational interests, and apparently can only be found if nonlinear models are employed. Another issue that deserves more attention in research on personality-intelligence associations is the "general factor of personality", which may be an important confound in understanding these associations.

Finally, the concept of trait complexes as originally proposed by Ackerman and colleagues appears to be untenable (Ackerman, 1996; Ackerman & Beier, 2003a). Based on their predictive validity, the overlap among cognitive abilities, personality and interests is best conceived as being among continuous variables. One of the most promising approaches is to use the dimensions that underlie the RIASEC model to anchor cognitive abilities and personality traits (Anthoney & Armstrong, 2010; Armstrong et al., 2008). The factors of interests and cognitive abilities found in chapter 5 bore a strong resemblance to the People/Things and prestige dimensions found in the models of Armstrong and colleagues. Hence, integrative models may be best based around the structure of the RIASEC model of occupational interests, although further research is needed to compare this approach more directly to PPIK theory, and to examine its capacity to predict occupational outcomes.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*(2), 227-257. doi: 10.1016/s0160-2896(96)90016-1

Ackerman, P. L. (2000). Domain-Specific knowledge as the "dark matter" of adult Intelligence, Gf/Gc, personality and interest correlates. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 55*(2), P69-P84.

Ackerman, P. L., & Beier, M. E. (2003a). Intelligence, personality, and interests in the career choice process. *Journal of Career Assessment, 11*(2), 205-218.

Ackerman, P. L., & Beier, M. E. (2003b). Trait complexes, congitive investment and domain knowledge. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Perspectives on the psychology of abilities, competencies, and expertise* (pp. 1-30). New York: Cambridge University Press.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: evidence for overlapping traits. *Psychological bulletin, 121*(2), 219.

Ackerman, P. L., & Rolfus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and non-ability traits. *Psychology and Aging, 14*, 314-330.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute, 50*(1), 277-291.

Anthoney, S. F., & Armstrong, P. I. (2010). Individuals and environments: Linking ability and skill ratings with interests. *Journal of counseling psychology, 57*(1), 36.

Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology, 97*(6), 1142.

Armstrong, P. I., Day, S. X., McVay, J. P., & Rounds, J. (2008). Holland's RIASEC model as an integrative framework for individual differences. *Journal of counseling psychology, 55*(1), 1-18.

Armstrong, P. I., Su, R., & Rounds, J. (2011). Vocational interests: The road less traveled. *The Wiley-Blackwell Handbook of Individual Differences*, 608-631.

Ashton, M. C., & Lee, K. (2012). On Models of Personality Structure. *European Journal of Personality, 26*(4), 433-434.

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: do they exist? *Personality and Social Psychology Review, 13*(2), 79-91.

Austin, E. J., Deary, I. J., & Gibson, G. J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence, 25*(1), 49-70. doi: http://dx.doi.org/10.1016/S0160-2896(97)90007-6

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

Austin, J. T., & Hanisch, K. A. (1990). Occupational attainment as a function of abilities and interests: A longitudinal analysis using Project TALENT data. *Journal of Applied Psychology, 75*(1), 77-86.

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*(3), 335-344.

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

Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*(1), 45-74.

Bartholomew, D. (1987). *Latent variable models and factor analysis*. London, England: Griffin.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychol Rev, 116*(3), 567-579.

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22*(1), 121-130.

Bouchard, T. J. (1997). Experience producing drive theory: How genes drive experience and shape personality. *Acta Paediatrica, 86*(S422), 60-64.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press; US.

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

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological bulletin, 40*(3), 153.

Cattell, R. B. (1950). *Personality*. New York: McGraw-Hill.

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

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. New York: Elsevier Science Publishing Company, Inc.

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

Chamorro-Premuzic, T., & Furnham, A. (2006). Intellectual Competence and the Intelligent Personality: A Third Way in Differential Psychology. *Review of General Psychology, 10*(3), 251-267.

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 102*(2), 408.

Conn, S. R., & Rieke, M. L. (Eds.). (1994). *16PF Fifth Edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

Costa Jr, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual*: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1995). Primary traits of Eysenck's PEN system: Three-and five-factor solutions. *Journal of personality and social psychology, 69*(2), 308.

Cramer, A. O., Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., . . . Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality, 26*(4), 414-431.

Cureton, E. E. (1968). *A factor analysis of Project Talent tests and four other batteries*. Pittsburgh, PA: American Institutes for Research and School of Education, University of Pittsburgh.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual-differences model and its applications*: University of Minnesota Press Minneapolis, MN.

De Fruyt, F., & Mervielde, I. (1999). RIASEC types and Big Five traits as predictors of employment status and nature of employment. *Personnel Psychology, 52*(3), 701-727.

De Raad, B., & Barelds, D. P. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of personality and social psychology, 94*(2), 347.

De Raad, B., & Peabody, D. (2005). Cross-culturally recurrent personality factors: analyses of three factors. *European Journal of Personality, 19*(6), 451-474.

de Vries, R. E. (2011). No evidence for a General Factor of Personality in the HEXACO Personality Inventory. *Journal of Research in Personality, 45*(2), 229-232.

Deary, I. J. (2009). The trait approach to personality G. M. Philip J. Corr (Ed.) *The Cambridge handbook of personality psychology* (pp. 89-109).

Deng, C.-P., Armstrong, P. I., & Rounds, J. (2007). The fit of Holland's RIASEC model to US occupations. *Journal of Vocational Behavior, 71*(1), 1-22.

DeYoung, C. G. (2011). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 711-737). New York: Cambridge University Press.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of personality, 73*(4), 825-858.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of personality and social psychology, 93*(5), 880.

Dunkel, C. S. (2013). The general factor of personality and general intelligence: Evidence for substantial association. *Intelligence, 41*(5), 423-427.

Einarsdóttir, S., & Rounds, J. (2000). Application of three dimensions of vocational interests to the Strong Interest Inventory. *Journal of Vocational Behavior, 56*(3), 363-379.

Eysenck, H. J. (1939). Review of Thurstone's Primary Mental Abilities, 1938. *British Journal of Educational Psychology, 39*, 188-231.

Eysenck, H. J. (1994). Personality and intelligence: Psychometric and experimental approaches. In R. J. Sternberg & P. Ruzgis (Eds.), *Personality and intelligence* (pp. 3-31). New York: Cambridge University Press.

Eysenck, H. J., & White, P. O. (1964). Personality and the measurement of intelligence. *British Journal of Educational Measurement, 24*, 197-201.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). New Jersey: John Wiley & Sons, Inc.

Flanagan, J. C., Dailey, J. T., Shaycroft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. (1962). *Design for a study of American youth*. Boston: Houghton Mifflin.

Flanagan, J. C., Davis, F. C., Dailey, J. T., Shaycroft, M. F., Orr, D. B., Goldberg, I., & Neyman, C. A. (1964). *The American high-school student: The identification, development, and utilization of human talents*. Pittsburg: Project Talent Office, University of PIttsburgh.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology, 59*(6), 1216.

Gottfredson, L. S. (1986). Occupational aptitude patterns map: Development and implications for a theory of job aptitude requirements. *Journal of Vocational Behavior, 29*(2), 254-291.

Gottfredson, L. S. (2005). Applying Gottfredson's theory of circumscription and compromise in career guidance and counseling. In S. D. Brown & R. W. Lent (Eds.), *Carreer development and counseling: Putting theory and research to work* (pp. 71-100). New York: John Wiley.

Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*(3), 179-203. doi: Doi: 10.1016/0160-2896(84)90008-4

Gustafsson, J.-E. (1988). Hierarchical models of individual difference in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in* *psychological and educational measurement* (pp. 73-95). London: Lawrence Erlbaum Associates, Publishers.

Hakistian, A. R., & Catell, R. B. (1975). *The comprehensive ability battery*. Champaign, IL: Institute for Personality and Ability Testing.

Harring, J. R., Weiss, B. A., & Hsu, J. C. (2012). A comparison of methods for estimating quadratic effects in nonlinear structural equation models. *Psychological Methods, 17*(2), 193-214.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical science*, 297-310.

Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska symposium on motivation: Vol. 30. Personality: Current theory and research* (pp. 55-89). Lincoln: University of Nebraska Press.

Hogan, R., & Roberts, B. W. (2000). A socioanalytic perspective on person–environment interaction.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6*(1), 35.

Holland, J. L. (1973). *Making vocational choices: A theory of careers* (Vol. 37): Prentice-Hall Englewood Cliffs, NJ.

Holland, J. L. (1985). *Vocational Preference Inventory (VPI) manual: 1985 edition*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources

Hopwood, C. J., Wright, A. G. C., & Brent Donnellan, M. (2011). Evaluating the evidence for the general factor of personality across multiple inventories. *Journal of Research in Personality, 45*(5), 468-478. doi: http://dx.doi.org/10.1016/j.jrp.2011.06.002

Horn, J. L. (1998). A basis for research on age differences in cognitive capacities. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 29-73). Mahwah, NJ: Erlbaum.

Horn, J. L., & Blankson, N. (2005). Foundations for a better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 41-68).

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

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

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78*, 250-250.

Hunt, E. B. (2011). *Human Intelligence*. Cambridge: Cambridge University Press.

Irwing, P., Booth, T., Nyborg, H., & Rushton, J. P. (2012). Are g and the general factor of personality (GFP) correlated? *Intelligence, 40*(3), 296-305.

Jackson, D. N. (1977). *Manual for the Jackson Vocational Interest Survey*. Port Huron, MI: Research Psychologists Press.

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

Johnson, W., & Bouchard, T. J. (2009). Linking abilities, interests, and sex via latent class analysis. *Journal of Career Assessment, 17*(1), 3-38.

Johnson, W., & Bouchard, T. J., Jr. (2005a). Constructive replication of the visual-perceptual-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability. *Intelligence, 33*(4), 417-430.

Johnson, W., & Bouchard, T. J., Jr. (2005b). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*(4), 393-416.

Johnson, W., Bouchard, T. J., Jr., McGue, M., Segal, N. L., Tellegen, A., Keyes, M., & Gottesman, I. I. (2007). Genetic and environmental influences on the Verbal-Perceptual-Image Rotation (VPR) model of the structure of mental abilities in the Minnesota study of twins reared apart. *Intelligence, 35*(6), 542-562.

Johnson, W., Jung, R. E., Colom, R., & Haier, R. J. (2008). Cognitive abilities independent of IQ correlate with regional brain structure. *Intelligence, 36*(1), 18-28. doi: DOI: 10.1016/j.intell.2007.01.005

Johnson, W., Te Nijenhuis, J., & Bouchard, T. J., Jr. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence, 35*(1), 69-81. doi: http://dx.doi.org/10.1016/j.intell.2006.05.002

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621-652.

Kan, K.-K., Kievet, R. A., Dolan, C. V., & van der Maas, H. L. (2011). On the interpretation of the CHC factor Gc. *Intelligence, 35*(5), 292-302.

Kanfer, R., Wolf, M. B., Kantrowitz, T. M., & Ackerman, P. L. (2010). Ability and trait complex predictors of academic and job performance: A person–situation approach. *Applied Psychology, 59*(1), 40-69.

Kaufman, J. C. (2012). Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences, 22*, 118-122.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 35*, 108-127.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence, 36*(6), 502-525.

Kenny, D. A. (2011). Measuring model fit Retrieved February 15, 2012, from http://davidakenny.net/cm/fit.htm

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*(4), 457-474.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 77*(6), 1121.

Kuder, G. F., & Richardson, M. W. (1937). The theory of test reliability. *Psychometrika, 2*, 151-160.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: can one construct predict them all? *Journal of personality and social psychology, 86*(1), 148-161.

Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36*(5), 422-436. doi: http://dx.doi.org/10.1016/j.intell.2007.08.004

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15*(3), 155-163. doi: 10.1111/j.0956-7976.2004.01503003.x

Loehlin, J. C. (2011). Correlation between general factors for personality and cognitive skills in the National Merit twin sample. *Journal of Research in Personality, 45*(5), 504-507.

Lohnes, P. R. (1966). *Measuring adolescent personality: Project Talent five-year follow-up studies, interim report 1*. Pittsburgh, PA: University of Pittsburgh.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences:"Sinking shafts at a few critical points". *Annual review of psychology, 51*(1), 405-444.

Luciano, M., Leisser, R., Wright, M. J., & Martin, N. G. (2004). Personality, arousal theory and the relationship to cognitive ability as measured by inspection time and IQ. *Personality and Individual Differences, 37*(5), 1081-1089. doi: http://dx.doi.org/10.1016/j.paid.2003.11.016

Lykken, D. T., Bouchard, T., McGue, M., & Tellegen, A. (1993). Heritability of interests: a twin study. *Journal of Applied Psychology, 78*(4), 649.

Mackintosh, N. J. (2012). *IQ and human intelligence* (2nd ed.). New York, NY: Oxford University Press.

Major, J. T., Johnson, W., & Deary, I. J. (2012). Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and extended Gf–Gc models. *Intelligence, 40*(6), 543-559.

McCrae, R., & Costa, P. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of personality, 57*(1), 17.

McCrae, R. R. (2009). The five-factor model of personality traits: Consensus and controversy. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 148-161): Cambridge University Press.

McCrae, R. R., & Costa Jr, P. T. (2012). *Personality in adulthood: A five-factor theory perspective*: Guilford Press.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology, 49*(3), 710.

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of personality and social psychology, 88*(3), 547.

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

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-181). New York, NY: Guilford Press; US.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. doi: http://dx.doi.org/10.1016/j.intell.2008.08.004

McLarnon, M. J., & Carswell, J. J. (2012). The personality differentiation by intelligence hypothesis: A measurement invariance investigation. *Personality and Individual Differences*.

Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. In T. Teo & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications*. Rotterdam, The Netherlands: Sense Publishers.

Moutafi, J., Furnham, A., & Paltiel, L. (2005). Can personality factors predict intelligence? *Personality and Individual Differences, 38*, 1021-1033.

Myers, I. B., McCaulley, M. H., & Most, R. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*: Consulting Psychologists Press Palo Alto, CA.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology, 66*(6), 574.

O'Connor, B. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(2), 396-402.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in* *psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.

Pincombe, J. L., Luciano, M., Martin, N. G., & Wright, M. J. (2007). Heritability of NEO PI-R extraversion facets and their relationship with IQ. *Twin Research and Human Genetics, 10*(3), 462-469.

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior, 21*(3), 259-287.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology, 25*, 111-164.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103-124.

Randahl, G. J. (1991). A typological analysis of the relations between measured vocational interests and abilities. *Journal of Vocational Behavior, 38*(3), 333-350.

Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than g. *Intelligence, 32*(6), 621-652. doi: http://dx.doi.org/10.1016/j.intell.2004.07.006

Reeve, C. L., & Bonaccio, S. (2011). The Nature and Structure of "Intelligence". *The Wiley-Blackwell Handbook of Individual Differences*.

Reeve, C. L., & Hakel, M. D. (2000). Toward an understanding of adult intellectual development: Investigating within-individual convergence of interest and knowledge profiles. *Journal of Applied Psychology, 85*(6), 897.

Reeve, C. L., & Heggestad, E. D. (2004). Differential relations between general cognitive ability and interest-vocation fit. *Journal of Occupational and Organizational Psychology, 77*(3), 385-402. doi: http://dx.doi.org/10.1348/0963179041752673

Reeve, C. L., Meyer, R. D., & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence, 34*(4), 387-402.

Rolfus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*, 174-188.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (Vol. 2): McGraw-Hill New York.

Rounds, J., & Tracey, T. J. (1993). Prediger's dimensional representation of Holland's RIASEC circumplex. *Journal of Applied Psychology, 78*(6), 875.

Rushton, J. P., & Irwing, P. (2008). A General Factor of Personality (GFP) from two meta-analyses of the Big Five: and. *Personality and Individual Differences, 45*(7), 679-683.

Sak, U. (2004). A synthesis of research on psychological types of gifted adolescents. *The Journal of Secondary Gifted Education, 15*(2), 70-79.

Scarr, S. (1996). How people make their own environments: Implications for parents and policy makers. *Psychology, Public Policy, and Law, 2*(2), 204.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of personality and social psychology, 86*(1), 162-173.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333-343.

Shaycroft, M. F. (1967). *The high school years: Growth in cognitive skills*. Pittsburgh, PA: American Institutes for Research and School of Education, University of Pittsburgh.

Snow, R. E. (1963). Effects of learner characteristics in learning from instructional films. Unpublished doctoral thesis, Purdue University.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 13-59). New York: Freeman.

Soubelet, A., & Salthouse, T. A. (2011). Personality-cognition relations across adulthood. *Developmental Psychology, 47*(2), 303-310.

Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201-292.

Spearman, C. (1939). Thurstone's work reworked. *Journal of Educational Psychology, 30*(1-16).

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Pearson.

Thurstone, L. L. (1938). *Primary mental abilities. Psychometric Monographs, No.1* Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of The Vectors of the Mind*. Chicago, IL: University of Chicago Press.

Tracey, T. J., & Rounds, J. (1992). Evaluating the RIASEC circumplex using high-point codes. *Journal of Vocational Behavior, 41*(3), 295-311.

Tracey, T. J., & Rounds, J. (1996). The spherical representation of vocational interests. *Journal of Vocational Behavior, 48*(1), 3-41.

Tracey, T. J., & Rounds, J. B. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological bulletin, 113*(2), 229.

Tupes, E. C., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of personality, 60*(2), 225-251.

Undheim, J. O. (1981). On intelligence IV: Toward a restoration of general intelligence. *Scandinavian Journal of Psychology, 22*, 251-265.

Undheim, J. O., & Gustafsson, J.-E. (1987). The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research, 22*, 149-171.

Van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842-861.

Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology, 96*(1), 13.

Vernon, P. E. (1961). *The structure of human abilities*. London: Methuen.

Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist, 20*, 723-733.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817.

Waiyavutti, C., Johnson, W., & Deary, I. J. (2012). Do personality scale items function differently in people with high and low IQ? *Psychological Assessment, 24*(3), 545.

Wise, L. L., McLaughlin, D. H., & Steel, L. (1979). *The Project TALENT data hanbook, revised*. Palo Alto, CA: American Institutes for Research.

Wolf, M. B., & Ackerman, P. L. (2005). Extraversion and intelligence: A meta-analytic investigation. *Personality and Individual Differences, 39*(3), 531-542.

Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*, 1090-1104.

Zeidner, M. (1995). Personality trait corrrelates of intelligence. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence*. New York: Plenum Press.

Zeidner, M., & Shani-Zinovich, I. (2011). Do academically gifted and nongifted students differ on the Big-Five and adaptive status? Some recent data and conclusions. *Personality and Individual Differences, 51*(5), 566-570.

Appendix A: Supplemental tables for chapter 3

Table A1
First-order factor loadings for grade 10 males in the broad selection (VPR, CHC and Gf-Gc models)

Test Name	Factor					
	Inform-ation	English/ Math	Spatial/ Reasoning	Mech./ Science	Speed	Math
Vocabulary	.64/.65/.65			.30/.31/.31		
Literature	.83/.84/.84					
Music	.73/.73/.73					
Social Studies	.84/.84/.84					
Mathematics	.32/.21/.21					.61/.68/.68
Physical Science	.31/.32/.32			.47/.46/.46		.16/.15/.16
Biological Science	.42/.43/.43			.36/.36/.36		
Aeronautics and Space	.33/.36/.36			.47/.47/.47		
Electronics				.82/.83/.83		
Mechanics				.74/.73/.73		
Art	.76/.76/.76					
Law	.68/.68/.68					
Health	.58/.59/.59	.05/.04/.04		.11/.13/.13		
Bible	.67/.67/.67					
Theatre and Ballet	.71/.71/.71					
Miscellaneous	.71/.71/.71					
Memory for sentences		.29/.29/.29				
Memory for words	.15/.14/.14	.39/.39/.39				
Disguised words	.33/.32/.32	.32/.34/.33			.25/.25/.26	
Spelling		.65/.64/.64				
Capitalization		.70/.69/.69				
Punctuation		.83/.83/.83				
English usage		.70/.70/.70				
Effective expression		.62/.62/.62				
Word functions in sent.		.42/.68/.67	.06/.09/.10			.33/–/–
Reading comprehension	.53/.53/.54	.32/.31/.31	.10/.10/.10			
Creativity	.34/.36/.36	.16/.15/.15	.32/.30/.31			
Mechanical reasoning			.58/.57/.57	.33/.39/.39		
Visualization in 2D			.56/.56/.56		.23/.21/.20	
Visualization in 3D			.77/.78/.78			
Abstract reasoning		.32/.30/.30	.50/.51/.51			
Math 1		.33/–/–	.19/.16/.15			.37/.68/.68
Math 2		.13/–/–				.78/.87/.97
Arithmetic comp.		.26/–/–			.34/.34/.34	.28/.50/.50
Table reading					.72/.72/.73	
Clerical checking					.69/.69/.69	
Object inspection			.15/.15/.14		.59/.59/.59	

Note. Loadings are in the order of VPR model, CHC model and Gf-Gc model

Table A2

First-order factor loadings for grade 10 females in the broad selection (VPR, CHC and Gf-Gc models)

Test Name	Factor					
	Inform-ation	English/ Math	Spatial/ Reasoning	Mech./ Science	Speed	Math
Vocabulary	.58/.60/.60	.12/.13/.13		.23/.23/.23		
Literature	.83/.83/.83					
Music	.76/.76/.76					
Social Studies	.83/.83/.83					
Mathematics	.23/.12/.12					.67/.72/.72
Physical Science				.66/.57/.56		.22/.31/.31
Biological Science	.28/.35/.35			.45/.40/.40		
Aeronautics and Space	.16/.21/.21			.38/.35/.35		
Electronics				.59/.62/.62		
Mechanics			.17/.20/.20	.51/.50/.51		
Art	.77/.77/.77					
Law	.61/.61/.61					
Health	.47/.50/.50	.13/.10/.10		.11/.11/.11		
Bible	.62/.63/.63					
Theatre and Ballet	.73/.73/.73					
Miscellaneous	.67/.67/.67					
Memory for sentences		.39/.38/.38				
Memory for words	.20/.17/.17	.41/.44/.44				
Disguised words	.33/.31/.31	.37/.38/.38			.27/.27/.27	
Spelling		.66/.65/.66				
Capitalization		.70/.69/.69				
Punctuation		.85/.84/.85				
English usage		.69/.69/.69				
Effective expression	.09/.07/.07	.52/.53/.53				
Word functions in sent.		.43/.67/.66	.11/.14/.15			.33/–/–
Reading comprehension	.53/.53/.53	.31/.30/.30	.11/.12/.12			
Creativity	.42/.42/.41		.34/.35/.35			
Mechanical reasoning			.64/.64/.64	.19/.19/.20		
Visualization in 2D			.59/.58/.58		.20/.20/.18	
Visualization in 3D			.75/.74/74			
Abstract reasoning		.34/.32/.32	.49/.50/.50			
Math 1		.36/–/–	.21/.20/.20			.31/.63/.63
Math 2		.21/–/–				.69/.84/.84
Arithmetic comp.		.44/–/–			.29/.30/.31	.18/.57/.57
Table reading		.25/.24/.23			.67/.67/.68	
Clerical checking					.69/.69/.68	
Object inspection			.24/.24/.22		.63/.63/.62	

Note. Loadings are in the order of VPR model, CHC model and Gf-Gc model

Appendix B: Supplemental tables for chapter 4

Table B1

Standardized linear and quadratic effects of g on the raw personality scales (males)

Trait	Linear effect				Quadratic effect			
	Gr. 9	Gr. 10	Gr. 11	Gr.12	Gr. 9	Gr. 10	Gr. 11	Gr. 12
Sociability								
Beta	.155	.108	.066	.049	-.119	-.118	-.124	-.132
R^2	.024	.012	.004	.002	.021	.020	.023	.025
Calmness								
Beta	.264	.261	.248	.246	–	–	–	–
R^2	.070	.068	.062	.061	–	–	–	–
Vigor								
Beta	.233	.196	.165	.146	-.094	-.091	-.084	-.090
R^2	.054	.038	.027	.021	.012	.011	.009	.011
Social Sensitivity								
Beta	.202	.202	.197	.193	–	–	-.018	-.024
R^2	.041	.041	.039	.037	–	–	.000	.001
Tidiness								
Beta	.212	.188	.142	.114	-.058	-.052	-.061	-.068
R^2	.045	.035	.020	.013	.004	.004	.005	.006
Culture								
Beta	.161	.156	.130	.132	–	.023	.040	.048
R^2	.026	.024	.017	.017	–	.001	.003	.004
Self-Confidence								
Beta	.245	.218	.212	.222	–	–	–	.017
R^2	.060	.047	.045	.049	–	–	–	.001
Mature Personality								
Beta	.273	.255	.241	.230	.044	.051	.054	.044
R^2	.075	.065	.058	.053	.003	.005	.006	.004
Impulsiveness								
Beta	–	.032	.030	.044	–	–	–	–
R^2	–	.001	.001	.002	–	–	–	–
Leadership								
Beta	.060	.071	.093	.116	.084	.080	.073	.065
R^2	.004	.005	.009	.013	.010	.010	.008	.007

Effects greater than .015 are significant at $p < .001$, with no adjustment for multiple testing. Non-significant effects are not shown.

Table B2

Standardized linear and quadratic effects of g on the personality scales (females)

Trait	Linear effect				Quadratic effect			
	Gr. 9	Gr. 10	Gr. 11	Gr.12	Gr. 9	Gr. 10	Gr. 11	Gr. 12
Sociability								
Beta	.129	.075	.028	-.023	-.138	-.149	-.148	-.138
R^2	.017	.006	.001	.000	.030	.033	.033	.030
Calmness								
Beta	.226	.207	.201	.175	*-.028*	-.033	-.028	-.037
R^2	.051	.043	.040	.031	.001	.001	.001	.002
Vigor								
Beta	.221	.177	.146	.118	-.078	-.088	-.077	-.067
R^2	.049	.031	.021	.014	.008	.011	.009	.006
Social Sensitivity								
Beta	.237	.217	.209	.175	-.060	-.084	-.086	-.086
R^2	.056	.047	.044	.031	.006	.011	.011	.011
Tidiness								
Beta	.176	.121	.086	.032	-.085	-.099	-.103	-.116
R^2	.031	.015	.007	.001	.011	.014	.016	.020
Culture								
Beta	.254	.252	.247	.226	-.014	-.016	-.011	-.004
R^2	.065	.063	.061	.051	.000	.000	.000	.000
Self-Confidence								
Beta	.185	.165	.164	.169	-.007	-.003	.011	.013
R^2	.034	.027	.027	.029	.000	.000	.000	.000
Mature Personality								
Beta	.282	.264	.288	.269	.059	.055	.050	.036
R^2	.080	.069	.083	.072	.005	.006	.004	.003
Impulsiveness								
Beta	.071	.119	.113	.117	.040	.046	.021	.023
R^2	.005	.014	.013	.014	.002	.003	.001	.001
Leadership								
Beta	.074	.057	.072	.095	.052	.047	.050	.056
R^2	.005	.003	.005	.009	.004	.004	.004	.005

Effects greater than .015 are significant at $p < .001$, with no adjustment for multiple testing. Non-significant effects are not shown.

Table B3

AIC differences between null and full GAM models (grade 10 males/females)

Trait	AIC difference - males	AIC difference - females
Sociability	1451.7	1927.4
Calmness	371.5	13.2
Vigor	841.1	406.6
Social sensitivity	16.1	303.0
Tidiness	258.2	969.0
Culture	522.8	275.2
Self-confidence	502.8	116.8
Mature personality	805.8	1781.1
Impulsiveness	268.3	456.5
Leadership	1145.0	1132.3

Note: the AIC for null model in males was 134924.2, and 135009.3 in females.

Appendix C: Supplemental tables for chapter 5

Table C1. Odds ratios of abilities and interests predicting job categories (grade 11 males).

Predictor	Job Category											
	Science	Med.	Business	Teach.	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
g	2.64	2.97	1.38	1.34	3.50	1.32	1.33		.56	.66	.50	
Spatial	1.38	1.35		.73		1.33					1.26	
English		1.43				1.19	.79			1.25		
Speed			1.23		.77						1.24	
Math	1.39					1.33	.84				1.17	
I: Trades	.73	.70	.81	.78	.46				1.65		1.49	
I: Politics			1.29	.72	1.43					.78		
I: Science	1.60	1.56		.78	.81	1.31	1.15		1.50			
I: Clerical	1.15			1.16	.82	.65	1.27		.82			
I: Medicine		2.27			.82	.77						
I: Arts	.75	1.21	.82	1.55		1.44		.70	1.17		.81	1.18
I: Teaching				1.74	1.35	.76			.70	1.27		
I: Military	1.25					.77		1.19				
I: Sales	.61	.77	1.18		1.66	1.71	.68		.78	1.54		
I: Arch.	1.54	.57	1.22	.79		1.29	1.28		.82			
F1: Trades	.17	.11	.55	.70	.08	.55	.81		3.05	1.71	3.71	2.91
F2: People	.23		1.16	2.66	2.55	.58	.56		.48	1.70	.76	
F3: Things	19.58	5.67	1.56	.61	2.62	3.70	2.31	.82	.80	.40	.53	.34
R^2: full	.38	.49	.10	.20	.51	.20	.09	.03	.24	.13	.23	.02
R^2: factors	.53	.51	.09	.18	.61	.17	.09	.01	.28	.10	.27	.21

Note: Odds ratios between .86 and 1.14 not shown. R^2: full was for individual scales as predictors; R^2: factors was for the trait-complex factors.
Med. = Medicine, Mech. = Mechanical.

Table C2. Odds ratios of abilities and interests predicting job categories (grade 11 females).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
g	2.51	1.96	1.83	1.33	1.82	.64	.80		.32		.28	.52
Spatial	1.40	.72	.79		.83	1.41		1.33	1.90		.39	
English	1.70	.67	.83		.80	1.34	.80		1.40		.40	
Speed		.84	1.33			.56			1.63		.56	
Math		.79	.78				1.17		1.34			
I: Trades	.34	1.26	.82	.85			.55		1.61			1.27
I: Politics	1.42	1.20		1.37	1.21	.82		1.49	.62		.65	
I: Science	2.47				1.24	.84			.58	1.39	1.77	
I: Clerical	2.47	1.33		.49	.69	.47			1.25	1.41	.43	
I: Medicine	2.75	3.23	.76		.82		1.90		1.15	.79	1.62	
I: Arts		.82	1.37		1.53	1.15	1.23	1.25	1.88		1.50	
I: Teaching	1.67		.81	1.70	1.25	2.35	.82	.56			.34	.81
I: Military						1.44	.75				16.63	1.33
I: Sales	.34	.64		1.15	.65			1.33	.66		1.66	.84
F1: Clerical	.53	.10	.58	.29	.32	.46		1.18	2.56	1.52	1.26	1.90
F2: People	1.24		2.66	5.20	2.59	2.97	.85	1.22	.27	.68	.44	.40
F3: Things	1.74	5.67	.51	.52		.63		.83	1.56	1.25	4.06	1.77
R^2: full	.53	.36	.19	.25	.26	.30	.19	.13	.45	.06	.80	.18
R^2: factors	.18	.51	.11	.32	.29	.16	.01	.02	.24	.04	.22	.11

Note: Odds ratios between .86 and 1.14 not shown. R^2: full was for individual scales as predictors; R^2: factors was for the trait-complex factors. Med. = Medicine, Mech. = Mechanical.

Table C3. Odds ratios of latent classes in predicting job category (grade 11 males).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
Class 1	.56	.09	.47			.79	.68	1.17	1.95	1.14	1.92	
Class 2	.83	.41	.70		.22	.79		1.17	1.51		.80	
Class 3	.15	.76		3.24	3.66		1.26	.18	.80	2.03	.32	1.43
Class 5	.60	.37	1.25	2.29	2.21	1.18	1.21	.62		1.24	1.47	.75
R^2	.10	.13	.03	.06	.22	.01	.01	.10	.02	.02	.07	.01

Note: Reference class is class 4. Odds ratios between .86 and 1.14 not shown. Med. = Medicine, Mech. = Mechanical

Table C4. Odds ratios of latent classes in predicting job category (grade 11 females).

Predictor	Job Category											
	Science	Med.	Business	Teaching	Humanities	Fine Arts	Technical	Sales	Mech.	Clerical	Construc.	Labour
Class 1	.11	.35	.67	.73	.53	1.53					n/a	1.21
Class 3	1.70	.73		2.11	1.89	1.52	1.27	1.30	.51	1.20	n/a	.73
Class 4	.58	2.60	1.24	2.20	5.29	1.57	1.29	1.48	.19	.67	n/a	.54
Class 5	.26	1.33	.42		2.67	1.24	.52	.60	.73	1.27	n/a	1.35
Class 6	.72	3.00	.85	2.00	.70	2.90	.32	1.49	.24		n/a	1.64
R^2	.23	.12	.04	.05	.15	.02	.04	.02	.09	.01	n/a	.03

Note: Reference class is class 2. Odds ratios between .86 and 1.14 not shown. Med. = Medicine, Mech. = Mechanical
[a] model did not converge.