A psychological study of the relationship between personality assessment for selection and change in self-perception

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

Master of Arts

in

Psychology

at Massey University, Manawatū,

New Zealand

Catherine Waldron

2018

Abstract

The use of personality assessments in employee selection is a growing global trend. However, there are numerous controversies in the literature regarding its utility and potential impacts. In addition, many of the ways in which personality assessments are being used in selection are neither aligned with research evidence, nor constrained by a code of ethics or extensive training.

The selection process itself is not focused on the wellbeing on job applicants. This is one possible reason why little to no research has investigated the potential effects of personality assessment for selection on job applicants' self-perception.

A review of the literature reveals several possible mechanisms for occupational personality assessment as an antecedent to change in self-perception, including positive and negative events, induced behaviour and biased scanning, and response construction.

This thesis investigates the relationship between the assessment of an individual's personality, and change to that individual's self-perception, across two independent sub-projects.

In Sub-project A, self-perception was assessed for a group of job applicants before and after completion of a personality assessment within a selection process.

In Sub-project B, self-perception was assessed for a group of students before and after completion of a personality assessment and receipt of a written results/feedback report.

The findings for both sub-projects demonstrate evidence of change to self-perception and support for personality assessment for selection as an antecedent to change in self-perception.

Possible explanations for these results are examined in relation to the mechanisms listed above. The limitations of the current studies are discussed and avenues for future research are recommended.

Sub-projects A and B represent a unique contribution to the literature in relation to both personality assessment in organisational settings, and self-perception change.

Acknowledgements

A huge and heartfelt thank you to my supervisor, Dr Gus Haberman. Your patience, encouragement and tireless support have been invaluable and this thesis could not have been completed without you. It has been my pleasure and privilege to work with you over the long course of this project.

Thanks to my amazing partner, Erik, who has picked up the slack and the pieces as needed and without complaint.

To all my family and friends who have listened sympathetically, proofread chapters, and chosen their words carefully – thank you.

My sincere gratitude also to SHL New Zealand Ltd, who generously sponsored my second sub-project by providing assessments and the time of their excellent and supportive staff.

Finally, it is appropriate to thankfully acknowledge the support of the Massey University Human Ethics Committee: Southern B. They examined and approved the research that was originally conceived as a genuine longitudinal investigation with considerably larger design, more participant groups, several ethical concomitants, and potential issues (Southern B, Application 13/88).

Table of Contents

Note: A list of references can be found in Appendix 1.

List of Figures

List of Tables

1 Introduction

"From an ethical perspective, organizations should be concerned with the effects of selection procedures on the psychological well-being of applicants" (Gilliland, 1993, p. 695).

1.1 The research context

Personality assessment for selection can be defined as the completion of one or more personality assessments, by a job applicant, for the purpose of their personality assessment results being used by a hiring organisation to inform decision-making during a selection process.

How the applicant responds to the items on a personality assessment and how those responses are subsequently interpreted[1] are factors that can influence the likelihood that the applicant will be offered the role for which they have applied (and/or other, perhaps future roles within the same organisation).

Personality assessments are also used in other occupational contexts, as well as clinical settings, but their use in personnel selection (Hough, 1998) is the focus of this thesis.

The following sections offer five perspectives on the current state of personality assessment for selection.

1.1.1 The use of personality assessment for selection is widespread and increasing

Personality assessment within the workplace is a billion-dollar, global industry, and the increasing use of personality assessments in selection is a documented worldwide trend (Gatewood, Feild, & Barrick, 2016; Highhouse, Doverspike, & Guion, 2016; The Economist, 2013; Weber, 2015).

In CEB's online survey of 1,406 Human Resources professionals from around the world (including New Zealand), 62% of respondents said that personality assessments were used in selection at their organisation (Kantrowitz, 2014).

In a survey of selection practices across 20 countries, New Zealand ranked fifth highest for use of personality assessments in selection (Ryan, McFarland, Baron, & Page, 1999).

[1] Interpretations are made both by the assessment tool's scoring methodology which converts responses into results, and by the decision-makers who receive and extrapolate from those results.

Taylor, Mills and O'Driscoll (1993) surveyed 99 organisations and 30 management consulting firms in New Zealand. More than 31% of the organisations and more than 67% of the consulting firms used personality assessments in selection. The authors concluded that selection methods and tests utilised in New Zealand organisations resembled overseas distributions, including frequent use of personality assessments. In a follow-up study, Taylor, Keelty and McDonnell (2002) surveyed 100 organisations and 30 recruitment consultancies in New Zealand. In this survey, 46% of responding organisations and 89% of consultancies reported using personality assessments in selection.

An Official Information Act request by the Public Service Association in 2013 revealed that personality assessments are widely used in the New Zealand Public Sector, including in selection. In 2012, government departments spent $1.5 million on psychometric testing including personality assessments (Public Service Association, 2013a, 2013b).

1.1.2 The use of personality assessment for selection is controversial

For the purposes of this thesis, the utility of personality assessment for selection is taken as a premise and is not disputed (Hough & Oswald, 2008).

However, reported judgements regarding this utility are multi-polar, and rely on a diverse range of rational and empirical results. The literature analysing the validity and benefits of personality assessment for selection has provided conflicting conclusions and the issue remains contentious (Carless, 2009; Ferguson & Lievens, 2017; Ion & Iliescu, 2017; Morgeson et al., 2007; Penney, David, & Witt, 2011; Prien, Schippmann, & Prien, 2003; Shaffer & Postlethwaite, 2012; Tett & Christiansen, 2007). In addition, cultural construals of personality are complex, and careful consideration is advised with regard to cross-cultural comparisons (Marsella & Leong, 1995).

Personality assessments are also often viewed negatively by job applicants (Ambrose & Rosse, 2003; Judge, Erez, & Bono, 1998; Schmidt, Oh, & Shaffer, 2016). While, in principle, most if not all personality assessments invade privacy, some are considered to be unnecessarily invasive (Camara & Merenda, 2000; J. W. Jones, 1991; Kravitz, Stinson, & Chavez, 1996).

In New Zealand, the 2013 disclosure of government spending on psychometric testing received significant media attention. In that same year, the New Zealand Employment Court ruled against organisational use of personality assessment results to dismiss an employee during a restructure. The judge stated that the personality assessment results were irrelevant criteria in this instance. (Employment Court Christchurch, 2013; Public Service Association, 2013b). A

similar case is now pending in relation to the Inland Revenue Department's 2017 restructuring attempt (Public Service Association, 2017).

1.1.3 The ways in which personality assessments are being used in selection do not align with research evidence

Many empirical studies have demonstrated the criterion-related validity of personality assessments via the pairing of personality traits with criterion variables of job performance (Highhouse et al., 2016).

Following the logic of criterion validation, studies that advocate the predictive power of personality assessments for future job performance report on criterion-related validity coefficients. Specifically, these are correlations between personality assessment scores for particular traits (such as the Five-factor Model trait Conscientiousness) and quantitative indicators of job performance on particular jobs (Bartram, 2005; Black, 2000; Schmidt et al., 2016; Tett & Christiansen, 2007).

In writing about the evidence-based best-practice use of personality assessments in selection, Carless provides the following recommendations pertaining to the development of indicators of job performance as well as the use of personality assessment scores:

> (1) conducting a job analysis prior to selection or using validity generalization research to determine the constructs assessed by psychological tests; (2) it is not advisable to report numeric ability test scores to managers nor should full personality profiles be reported; (3) psychological test information should not be previewed prior to an interview; (4) psychological test data should be numerically combined with interview data; and (5) hiring managers should be encouraged to make their own decision about the suitability of an applicant. (2009, pp. 2526-2527)

However, studies investigating whether organisations (in New Zealand and elsewhere) adhere to best practice in their use of selection methods have found several gaps between research and application (Harris, Toulson, & Livingston, 1996; E. K. Johnson, 2000; Lawler, 2007; Rynes, Colbert, & Brown, 2002; Sanders, van Riemsdijk, & Groen, 2008; Taylor et al., 2002; Taylor et al., 1993).

Taylor, Mills and O'Driscoll (1993) reported that formal job analysis rarely preceded the use of personality assessment for selection or the choice of psychometric tool. Taylor et al (2002) found that only two of 45 organisations using personality assessments (4%) reported a focus

on specific job-related competencies in orienting the use of personality assessment results. Twenty-six of the 45 (58%) 'viewed results as a whole'.

These practices directly contrast with Carless's first and second recommendations (quoted above).

Anecdotally, some New Zealand organisations use personality assessment results in an impromptu fashion to predict the 'fit' between an applicant and the role, team, manager, or entire organisation (Kristof, 1996; Kristof-Brown, Zimmerman, & Johnson, 2005; Lauver & Kristof-Brown, 2001). This approach typically involves the selection panel viewing the complete personality assessment report for each applicant either before or after the interview, again deviating from the best practice guidelines.

1.1.4 The ways in which personality assessments are being used in selection are not constrained by a code of ethics or extensive training

Robertson and Smith (1989) point out that selection processes are not psychologically neutral but represent interventions in candidates' careers that can have consequences for wellbeing.

Yet in many organisational settings, personality assessments may be selected and administered by employees whose exposure to psychology and psychometrics training is limited to attendance at a short accreditation course provided by the assessment vendor. At least three assessment vendors in New Zealand currently offer accreditation courses that require one week or less of training, with additional digital provisions in some cases. Such courses might be more appropriate if Masters-level education in Psychology was a pre-requisite.

In their discussion of professional and ethical issues related to assessment for selection, Prien et al. (2003) emphasise the importance of the distinction between trained support staff and professional psychological 'assessors', where the expertise of the latter comes from tertiary education, training and experience.

The New Zealand Council of Education Research requires that users of personality assessments have completed advanced courses in psychometric assessment as well as personality or abnormal psychology, and notes that some for some assessments, specific training in that assessment is also required (New Zealand Council of Education Research, 2017).

Taylor et al. (1993) articulated concern regarding the widespread use in New Zealand of personality assessments that can be conducted in the absence of psychologists and/or university graduates to administer them.

While psychologists in New Zealand are extensively trained and are bound by a code of ethical practice (including guidelines for the use of psychometric tests; Code of Ethics Review Group, 2012; New Zealand Psychologists Board, 2015), there are no such controls in place for employees of hiring organisations.

Arguably, mitigating any potential negative impact to an individual as a result of a personality assessment is likely to be beyond the scope of an organisational administrator's role and capability. The same can be said for possessing (and being able to provide to applicants) an objective view of the strengths and limitations of personality assessments, and psychometrics in general.

1.1.5 The selection process, including personality assessment for selection, is not focused on applicant wellbeing

At the simplest level, vendors who develop and sell personality assessments for selection are motivated by commercialism[2]. Therefore, their focus will be on satisfying the needs of their customers (hiring organisations). How the selection process affects an applicant is likely to be of concern only in so far as it impacts on the vendor's reputation or brand.

Similarly, organisations that purchase and utilise personality assessments for selection are motivated by the desire to improve organisational outcomes (such as decrease in turnover costs) through recruitment of the most suitable candidate. How the selection process affects an applicant is likely to be of concern only in so far as it impacts on the organisation.

While surely neither vendors nor organisations have any desire to cause harm, applicant wellbeing is not the focus of the transaction.

1.2 The research problem

1.2.1 Potential impacts of personality assessment for selection on applicants have been ignored

[2] Goldberg et al (2006) refer to four scientific activities that could contribute to personality assessment improvements that are disallowed or discouraged by test publishers due to their commercial interests.

The perspectives above serve to illustrate a situation in which a widespread and growing practice is occurring independently of academic recommendations and ethical constraints, despite ongoing controversy regarding its utility. Further, perhaps due to the fact that the selection process is not focused on applicant wellbeing, the potential impacts of personality assessment for selection on job applicants have been largely ignored.

Within organisational psychology, applicant reactions research investigates job applicants' perceptions of selection processes as well as the antecedents and consequences of these perceptions (Imus & Ryan, 2005). However, the predominant focus of this area has been on potential negative outcomes for hiring organisations, such as decreased likelihood of applicants applying for a future role at the organisation (Ambrose & Rosse, 2003; Ryan & Ployhart, 2000). At least one prominent textbook refers only to organisation-related outcomes as justification for conducting applicant reactions research at all, with no mention of potential impacts on applicants themselves (McFarland, 2013).

While there are studies that have investigated applicant-related outcomes associated with selection, these have typically focused on how selection processes and procedures (including personality assessments) influence perceptions of fairness, which are then linked with particular outcomes, such as change in self-efficacy (Ployhart & Ryan, 1997). A second area has targeted how applicant perceptions affect performance within the selection process itself (Imus & Ryan, 2005).

No studies have surfaced that have specifically studied the direct consequences of personality assessment for selection on applicant self-perception or any other self-related construct.

1.2.2 Personality assessment for selection as antecedent to change in self-perception

The research literature provides evidence of a number of antecedents to change in self-perception, as well as potential effects of personality assessment.

Three areas within the literature hold potential for fruitful comparisons with personality assessment for selection as an antecedent to change in self-perception.

1) The theory of response construction (Feldman & Lynch, 1988; Peterson, 2005; Simmons, Bickart, & Lynch Jr, 1993) proposes that when an individual is required to answer a question about themselves and a previous response is not immediately available in memory, a new response is 'constructed' from available data, including the context within which the question has been asked. Completing a personality assessment may cause an individual to ask

themselves questions they have never considered before, and via response construction, a changed self-perception may emerge.

2) Stake, Huff and Zand (1995) demonstrated that anticipated positive and negative events could produce changes to anticipated self-perception. Personality assessment for selection provides multiple opportunities for an applicant to experience a positive or negative event, including completion of the assessment itself, receiving results/feedback, and success or failure within a selection process in which personality assessment results were a factor.

3) In experimental social psychology, multiple authors (encompassing multiple theoretical viewpoints) have demonstrated that induced behaviour can result in a change to self-perception that aligns with the induced behaviour (Fazio, Effrein, & Falender, 1981; E. E. Jones, Rhodewalt, Berglas, & Skelton, 1981; Tice, 1992). On the basis that completion of a personality assessment is a self-presentation (J. A. Johnson, 1981), the demand characteristics of the selection process may produce an 'induced' presentation that subsequently changes self-perception in line with the presentation.

1.3 The current research

1.3.1 Overview

The objective of the current research was to explore the broad hypothesis that completing a personality assessment and/or receiving results/feedback from that assessment could cause a change to self-perception.

Two sub-projects were undertaken, one involving a group of job applicants (Sub-project A) and the other involving a group of students (Sub-project B). In Sub-project A, self-perception was assessed before and after completion of a personality assessment within a selection process. In Sub-project B, self-perception was assessed before and after completion of a personality assessment and receipt of a written results/feedback report.

The findings for both sub-projects demonstrated change in self-perception after completion of the personality assessment (and receipt of results/feedback for Sub-project B), though the direction of change and the implicated components of self-perception contrasted across the two studies.

1.3.2 The structure of this thesis

This introduction chapter presents the research context, research problem and a brief overview of the current research.

Chapter two provides a review of the literature in relation to personality assessment, selection, and self-perception, with an emphasis on the five perspectives described above.

The third chapter provides operationalisation of the research objective, detail regarding the original and eventual research designs, and a summary of the resulting hypotheses for the two sub-projects.

Chapters four and five detail the research methods and results for Sub-projects A and B respectively.

The final chapter discusses and critiques the findings as a whole and examines their contribution to the wider literature. A concluding review addresses the limitations of the current studies and the considerations they raise for future research.

2 Literature review

2.1 Personality assessment

As discussed in the Introduction chapter, personality assessment for organisational purposes is a highly profitable, growing industry.

An indicative example is the Myers-Briggs Type Indicator, which is typology-based rather than a mainstream tool; this instrument alone is used by approximately 80% of Fortune 100 companies, and has been completed by millions of individuals each year since its publication in 1962 (Bajic, 2015; The Myers and Briggs Foundation, 2016).

While it is not known how many organisations in New Zealand use personality assessments in selection, research suggests that the breadth of usage and the application of results by New Zealand organisations are similar to elsewhere (Taylor et al., 1993).

The use of personality assessment in the human resources sector is centred within a nexus of well-documented controversies (Guion, 2011; Hough & Connelly, 2013; Hough & Oswald, 2000; Hough & Oswald, 2008), despite a large quantity of published research, and the popularity of the practice throughout the industry.

These controversies exist on several levels. On the theoretical plane, they relate to specific theories and conceptual models of personality (Burger, 2015; Larsen, Buss, & Wismeijer, 2013). Further dissension relates to theoretical assumptions underlying personality assessment including measurement issues (Michell, 1997) and the status of psychological constructs in general (Valentine, 1992). On the practical plane, professional, ethical, moral, and societal concomitants of personality testing procedures remain disputed. Personality assessments tend to be viewed unfavourably by job applicants, and their utility for selection is a topic of passionate debate.

The following sections provide a review of these areas of contention.

2.1.1 Theorising personality

Definitions of personality

Alongside other psychological constructs, personality is a commonly invoked predictor of behaviour (Burger, 2015; Carducci, 2015). Personality and its constituents are used in several modes of academic explanation for interindividual differences in explicit behaviour across

otherwise similar settings. In particular, personality traits as a specific type of personality variable are commonly employed to identify longer-term and situation-aspecific propensities or tendencies of individual behaviour (Mischel, Shoda, & Ayduk, 2008).

Given multiple theoretical approaches to human personality, it is perhaps unsurprising that both 'personality' per se and named traits are often exploited without accurate, consistent conceptual definition, or indeed without any acceptable definition (Allport, 1937; Bergner, 2017; Buss, 2008; Christiansen & Tett, 2013; R. Hogan & Blickle, 2013).

When a characterisation of a personality construct (Slaney, 2017) is suggested rather than precisely stated, divergent ontological and epistemological assumptions that have implications for theorising and measurement remain tacit or even unclear[3]. This insufficient clarity of definition is encountered particularly often when organisational psychologists and human resources experts refer to personality traits.

Major theoretical approaches

Personality research and measurement is typically split into two perspectives: the study of law-like differences among individuals (nomothetic approach), and the study of individual persons as unique integrated 'wholes', with non-comparable features (idiographic approach) (Allport, 1937; Barenbaum & Winter, 2008; Buss, 2008; Cervone & Mischel, 2002; McAdams, 2009; Shadel, 2004).

The nomothetic view has become globally prevalent in describing individual differences (Grice, 2015)[4] as well as in mainstream personality assessment, including organisational contexts (Buss, 2008; Lamiell, 2013). Idiographically-based theories and measuring instruments do exist (e.g. Repertory Grid) but are rarely utilised in the setting targeted by this thesis.

The four categories of personality theory covered below are those most common to the use of personality assessments in organisations. They are not exclusively associated with either the

[3] This situation is typical of Dekker and Hollnagel's (2004) discussion of the use of concept labels (or constructs) that are "intuitively meaningful in the sense that everyone associates something with them, so they feel that they understand them" (p. 79). This applies equally to the construct of self, and by extension to constructs predicated on self. Dekker and Hollnagel discuss how failure to specify what is actually meant by these labels can lead to and perpetuate a lack of clarity in published research. Some authors argue that is a problem for constructs in psychology in general (Michell, 2013; Slaney & Racine, 2013).

[4] Research into individual differences also tends to be associated with quantitative methodologies. Michell (Michell, 1997, 2000, 2004, 2008, 2012, 2013) has written extensively on the bias within psychology towards quantitative analyses.

nomothetic or idiographic approach, although proponents of trait-focused theories universally adopt nomethesis.

Trait theories

Within mainstream differential psychology, psychometrics, and human resources applications, the dominant personality model is the trait model (R. Hogan & Blickle, 2013). The impetus for theorising on traits can be traced back to Gordon Allport; for a modern conceptual analysis of the features of traits as a particular type of construct, see Paunonen & Hong (2015).

Theories based on the trait model propose that personality is comprised of traits, which are typically defined as "consistent patterns of behavior" (Barenbaum & Winter, 2008, p. 11). However, "they also may refer to postulated physiological and/or psychological attributes that generate consistent forms of behavior" (Cervone, 1991, p. 372).

McAdams (2009) outlines four positions held by theorists regarding the ontological nature of traits: traits as neurophysiological substrates, behavioural dispositions, act frequencies, and linguistic categories. Hogan and Foster (2016) argue that trait theory has a logical problem in that it is either defining traits simultaneously in multiple incompatible ways, or creating a tautology via traits as behavioural patterns being called upon to explain and predict behavioural patterns.

Aside from ontology, there is consensus that traits are stable over time, bipolar terms (i.e. on continua), additive and independent, and representative of broad individual differences in socio-emotional functioning (McAdams, 2009, p. 109).

Because traits are linked to behaviour, trait theorists assert that measurement of traits can provide predictions regarding future behaviour (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001). Self-report assessments of personality are held to measure true personality (to the extent that individuals respond truthfully to test items).

Type theories

Typologies assume that personality is a fixed pattern of dispositions that a person 'has'; and that this pattern can be fruitfully used to describe a large number of individuals. Unlike trait theories, personality types are identified in relation to discrete categories rather than along continua. In major typologies such as Carl G. Jung's, the number of types is small, and there is an assumption that each person in a population can be assigned to one type.

Despite the dominance of the trait model in differential psychology and psychometric research, one of the most widely used personality assessments is the Myers-Briggs Type Indicator (MBTI; Gerras & Wong, 2016; Neukrug, 2015; Reynierse, 2013). The MBTI is based on Jungian typology and the system of four dispositions (Jung, 1923). These are not dispositions in the sense of mainstream Anglo-Saxon psychological theory but rather flexible, bipolar preferences.

"A fundamental feature of Jung's theory – and therefore the construction and accurate interpretation of the MBTI instrument – is that it postulates qualitatively distinct categories rather than more familiar behavioural traits that vary along a continuum" (Quenk, 2009, p. 19).

Personality theory also recognises typological systems other than psychoanalytic (including Jungian); see Barenbaum and Winter (2008).

Socio-analytic theory

Socio-analytic theory (R. Hogan, 1982) is similar to trait theories in that it also uses trait terms, but it has a conflicting theoretical foundation regarding the ontology and epistemology of traits.

Socio-analytic theory differentiates between personality as reputation (what others believe about an individual's personality) and personality as identity (what the individual believes about their own personality).

Reputation is defined as a person's distinctive interpersonal style, including the kind of impression they make on others, while identity is defined as the causes or reasons for a person's style of behaviour or unique reputation (R. Hogan, 1982, p. 58).

Reputation, identity, and social skill (the ability to bring reputation in line with identity), are the three essential ways in which individuals differ from one another (while sharing common goals of getting along, getting ahead and finding meaning), and they are the units of analysis of personality (R. Hogan & Foster, 2016).

According to socio-analytic theory, trait terms refer to units of reputation (not identity), and it is reputation that is organised in line with the Five-factor Model, while little is known about identity (R. Hogan & Blickle, 2013).

This theory also entails that responses on personality assessments are interpreted as self-presentations rather than self-reports (R. Hogan & Foster, 2016).

Social cognitive theories

With the advent of radical information processing paradigms in cognitive psychology, theoretical constructs in personality have been increasingly linked with mental information processing and mental representations (Cervone & Pervin, 2013). Bandura's social learning theory harmonises with the main tenets of this theoretical shift, and the social learning approach has informed cognitive work on personality.

According to the social-cognitive perspective, behaviour is the result of interaction between the person and the social situation. The actor has agency (Bandura, 1999), and the interaction between person and environment is reciprocal, consistent with Bandura's original thesis of reciprocal determination (Cervone, Shadel, & Jencius, 2001).

When predicting overt behaviour, one must take the situation into account and not only the person. As in social cognition, predicting interpersonal behaviour proceeds through characterising mental representations that are the outcome of, and are further modified by, formally definable operations. The personality constellation of an individual is mapped to specific sets of representations (Cervone & Pervin, 2013).

Therefore, in both socio-analytic and social cognitive theories, measurable personality is something a person 'does', rather than something she/he possesses.

Five-factor Model

The Five-factor Model (FFM) is an extremely popular and well-regarded theory that describes the core structure of personality. It is often held to be responsible for the reawakening of interest and support for personality psychology (Barrick et al., 2001; Goldberg, 1993).

The FFM emerged from factor analysis based on the lexical hypothesis that important individual differences in personality will be reflected in common language. Through correlational analysis of attributions of a large number of common language personality-related adjectives, high level personality traits have been 'factored out' using statistical procedures (Lamiell, 2000).

Numerous investigations of this kind have provided support for five higher order personality traits collectively known as the Big Five, both from atheoretical factor analyses founded on the lexical hypothesis, and post-hoc factor analyses of theory-driven personality models (Bartram & Brown, 2005).

The FFM is primarily a nomothetic trait theory, however it is frequently suggested that other theories can be subsumed under its structure (Bartram & Brown, 2005; Cervone et al., 2001; McAdams, 2009; Reynierse, 2013). Findings in relation to the FFM are consistent with Hogan's socio-analytic theory[5], when defined as units of reputation, as described above.

The Big Five are typically labelled Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness (McAdams, 2009), though some disagreement remains over this.

Further, despite the widespread popularity of the FFM, there is controversy surrounding what individual differences data can legitimately reveal about individuals. This is discussed in more detail below.

2.1.2 Assumptions underlying the assessment of personality

Personality exists

The assessment of personality logically presupposes that personality exists (Hanson, 1993).

Maul (2013) describes how clarity regarding ontology should predate claims to measurement of a construct, and yet rarely does so for psychological attributes such as personality.

Boag uses a "realist account of test validity" to argue that logical problems within (trait) definitions of personality prevent a satisfactory account of the existence of personality. He refers to "long-standing arguments [that] propose that dispositional traits are descriptive summaries reified into within-subject attributes and then erroneously used to explain individual behaviour." (2015, p. 36).

This is the same line of argument employed by Hogan and Foster (2016) in their critique of the trait model (see above).

While these criticisms are primarily directed towards trait theories, Slaney (2017; Slaney & Garcia, 2015; Slaney & Racine, 2013) has written extensively on related issues with definitions of constructs in psychology (such as personality). She details how constructs as concepts have been conflated with the phenomena they are intended to represent, with flow-on effects including ambiguity in relation to ontology and epistemology.

[5] Socio-analytic theory also posits two higher order factors underlying the Big Five: successful socialisation (getting along) and personal growth and self-enhancement (getting ahead; J. Hogan & Holland, 2003).

Personality is 'quantitative' (and therefore measurable)

Though proponents of idiography may question quantification, most modern assessments are predicated on the tacit assumption that all or most aspects of human personality can be quantified, i.e. they exist in amounts that can be numerically represented. Mainstream personality tests, such as those used in selection, rely on the view that everything studied in relation to personality can be mapped to quantifiable attributes.

Joel Michell outlines three categories of structure pertaining to attributes (or constructs):

(1) *"classificatory attributes* (with heterogeneous differences between categories);
(2) *heterogeneous orders* (with heterogeneous differences between degrees); and
(3) *quantitative attributes* (with thoroughly homogeneous differences between magnitudes)" (2012, p. 2).

Michell has written extensively about the lack of evidence for (and inquiry into) the assumption that psychological constructs, such as personality, have quantitative structure (1997, 2000, 2004, 2008, 2012, 2013).

His writing clarifies an important distinction between linear transformations (i.e. quantitative representations of item responses as numerical values/numeric coding) and true interval-level data.

It is typical within psychometric tests (including personality assessments) to apply numerical values to test item responses (e.g. For Item *x*, Strongly Agree is scored as a 5, Agree as a 4, and so on), and to then conduct statistical procedures using these scores.

Putting aside important and often neglected considerations such as how individual participants interpret and respond to item response labels such as Strongly Agree (Grice, 2015), many of the statistical procedures commonly applied to assessment scores rely on the assumption that the scores represent interval-level data.

But interval-level data require that differences between scores are homogeneous, i.e., the difference between Disagree and Strongly Disagree is the same as the difference between Agree and Strongly Agree. And this can only result from measurement (in its true sense) of quantitative attributes.

On the basis of these arguments, not only is it premature to call personality assessment scores a 'measurement' of personality, but the statistical manipulations commonly performed on such scores may also be unjustified. Krause (2013) points out that the amount of distortion

provided by the assumption of interval-level data can only be discovered when there are results predicated on the assumption of ordinal-level data with which to compare them.

Personality is stable

Within the trait model in particular, it is posited that personality is inherently stable over time and across situations (Roberts & DelVecchio, 2000).

Therefore personality assessment provides a measure of something stable, which can be used to make inferences about the future. This has clear significance for the use of personality assessments in selection, where future job performance is predicted on the basis of current personality.

Traditionally, psychology has differentiated between states, which are temporary, and traits, which are static (McAdams, 2009).

However a number of studies have demonstrated that personality traits may be less stable than has been thought (Boyce, Wood, & Powdthavee, 2013; Klimstra, Bleidorn, Asendorpf, van Aken, & Denissen, 2013; Martin, Oades, & Caputi, 2013).

For example,

- Specht et al. (2014) reported that over a lifetime, there are periods of comparatively high stability (middle adulthood) and periods of comparatively strong changes (young adulthood and old age) in personality.
- Sheldon, Ryan, Rawsthorne and Ilardi (1997) demonstrated that people show varying levels of Big Five personality traits depending on their assumed role, such as employee, friend, student, and romantic partner, and that this varies as a function of 'felt authenticity'.
- Hudson & Fraley (2015) found that subjects could intentionally cause change to their personalities.

The assumption of stability within personality is also linked to the reliability of personality assessment tools. Sackett and Walmsley (2014) report on three categories of evidence for personality change, including test-retest correlations for personality scores over time.

Threats to the stability of personality have implications not only for the reliability of personality assessment results, but also for how these results are applied and communicated.

Population statistics provide meaningful data about individuals

Multiple authors have signaled a significant issue in relation to what can legitimately be said about individuals on the basis of population statistics, with specific reference to the use of data derived from the Five-factor Model.

Personality assessments based on the FFM are frequently used to assign to individuals a personal rating or score for each of the Big Five traits. However, the Big Five are factors that have been extracted from analyses of interindividual differences in personality, rather than intra-individual investigations of personality.

Lamiell has written extensively regarding the pitfalls involved in drawing conclusions about individuals on the basis of research into differences between individuals (2000, 2007, 2010, 2013). Molenaar and Campbell (2009) propose that classical mathematical-statistical theorems show that interindividual analyses can only provide information about the population, and therefore cannot be applied at the level of the individual.

In addition, Grice (2015) points out that FFM factors do not regularly emerge from analyses of data from individuals. "There is a genuine and potentially hazardous disconnect, then, between conclusions drawn from between-persons aggregate statistics and statements or theories meant to offer insight into the psychology of individual persons." (Grice, 2015, p. 1).

2.1.3 Personality assessment in organisational contexts

Personality as a predictor of job performance

The research evidence in support of the application of personality assessments in selection and other organisational contexts is centred on personality as a predictor of job performance. This research primarily relies on FFM personality scores (Barrick et al., 2001; N. Schmitt, 2014).

"The cumulated evidence in the past fourteen years has demonstrated that personality variables have a strong impact on job performance and other relevant criteria for organizations (e.g., absenteeism, counterproductive behaviours, promotions, turnover) and, consequently, are now accepted as useful tools for personnel selection and included in many models of job performance" (Salgado & De Fruyt, 2005, pp. 191-192).

The mechanism by which personality affects performance is not known but is assumed to be linked to motivation (Barrick et al., 2001; Penney et al., 2011).

“The most widely-accepted theories regarding the relationship between personality and job performance focus on work motivation as the key mediating mechanism (Barrick & Mount, 2005; Hogan, 1996; Hogan & Holland, 2003; Kanfer & Ackerman, 2000). Although motivation has been defined in many ways, the most commonly cited theories focus on the cognitive processes underlying goal-setting, defined as the “arousal, direction, intensity and persistence of voluntary actions that are goal directed” (Mitchell, 1997, p. 60)“ (Penney et al., 2011, p. 298).

However, there is disagreement in the literature that the evidence in support of personality as a predictor of job performance is sufficient to justify its use, with detractors often referring to low validities (Morgeson et al., 2007; Smith & George, 1994).

Another substantial source of dispute is response distortion, or faking.

With most personality inventories utilising self-report, the veracity of responses and related potential impact on scores has provoked widespread empirical work (Hough & Oswald, 2008; Morgeson et al., 2007), and the literature in this area is contentious.

Authors disagree on the causes of response distortion, real-time processes during testing, the practical impact of faked responses, and ways of revealing and reducing distortion. One group of authors describe the impact as minimal and not warranting scientific concern (N. Schmitt & Oswald, 2006); others contend that the extent of faking justifies the cessation of the use of self-report procedures altogether (Hough & Oswald, 2008.

Response distortion is also complicated by internally heterogeneous categories. Responses may be distorted because of less conscious or less intentional processes (e.g. deficient interpretation of items or lack of self-knowledge), or because test-takers deliberately misrepresent themselves. Specific aspects such as response sets and response styles have also prompted investigation.

Rothstein and Goffin (2006) reviewed current methodologies for controlling response distortion via assessment design. They advocate the inclusion of faking warnings in test instructions, and ipsative test designs (e.g. forced-choice items and scoring) as effective options.

Conversely, Hogan and others propose that a test-taker who answers positively in order to meet the demands of the situation is showing adaptive behaviour that should be valued (R. Hogan & Foster, 2016). “An organization should be indifferent to whether a given pattern of valued behavior is or is not consistent with a person’s underlying disposition as long as the valued behavior is exhibited.” (Sackett & Walmsley, 2014, p. 540).

Approaches to test development

As noted above, elements of test design are implicated in disputes regarding the utility of personality assessment for selection. The following sub-sections briefly summarise three salient distinctions between assessment tools: data source, test and item design (e.g. format of test items), and contextualisation.

Self-report versus observer ratings

In assessing an individual's personality, data can be obtained via questions posed to the individual (known as self-report), or to others who have observed the individual (known as observer ratings; Gatewood et al., 2016).

Self-report is by far the more frequently used method of data collection in assessing personality. "Personality measurement is almost synonymous with standardized self-report questionnaires" (Hough & Dilchert, 2010, p. 301).

Funder reports that, "according to most research, the way people describe themselves by and large matches the way they are described by others" (2010, p. 25).

However, the reliability of self-report personality assessments is controversial in settings where the outcome of the personality assessment has bearing on the individual's future, such as personality assessment for selection, due to concern about applicant faking.

"Personality constructs certainly have value in understanding work behavior, but future research should focus on finding alternatives to self-report personality measures." (Morgeson et al., 2007, p. 721).

Test item format and scoring

Within standardised personality assessments, the two most frequently used formats for test items are Likert rating-scales and forced choice (Bartram, 2007). Both formats have traditionally relied on Classical Test Theory.

Likert-type scales are used regularly and there is a wide knowledge base regarding their properties, however there are concerns that they are vulnerable to response distortion (Hartley, 2014; O'Neill et al., 2017; Rothstein & Goffin, 2006).

Alternatively, forced choice response formats have been proposed as being more resistant to faking, but they are considered to provide ipsative results that are less useful (Ones, Dilchert, Viswesvaran, & Judge, 2007; Saville & Willson, 1991).

More recently, forced choice scoring using computer-adaptive testing based on Item Response Theory has demonstrated normative results (Foster, Min, & Zickar, 2017; Gatewood et al., 2016; Joubert, Inceoglu, Bartram, Dowdeswell, & Lin, 2015).

Contextualised versus non-contextualised tools

A third categorisation separates contextualised and non-contextualised assessments. For the current purpose, 'contextualised' is a term qualifying tools that have been originally intended and developed for a specific context, such as human resources. Contextualised tools direct test-takers to respond to items with regard to the specified context.

Non-contextualised personality assessments are those that have been created for a wide scope of uses, rather than one selected context.

“Personality predictors used in personnel selection can be divided roughly into two categories. First, there are measures of normal adult personality. The initial purpose in the construction of these measures was the accurate description of individual differences in personality. That is, they were developed to provide broad descriptions of personality that could be used in a wide range of settings (….). The second category of personality measures used in personnel screening and selection can be referred to loosely as measures of personality at work. The initial purpose in construction of these measures was the accurate prediction of individual differences in work behaviours of interest” (Ones & Viswesvaran, 2001, p. 64).

This is an important distinction because contextualisation increases the point-to-point correspondence between the personality assessment and work-related information (Asher & Sciarrino, 1974), and, perhaps for this reason, contextualised tools show increased predictive validity in relation to job performance (Shaffer & Postlethwaite, 2012).

Contextualised assessments used in selection also have greater face validity for job applicants.

Specific tools

The following sub-sections provide a brief description of several well-known and widely used personality assessment tools. All of them are self-report inventories, and unless noted otherwise, they adopt a trait approach.

Non-contextualised tools

Minnesota Multiphasic Personality Inventory (MMPI)

The MMPI was developed in the 1940s by Ben-Porath and Tellegen as an assessment of psychopathology (Neukrug, 2015). It quickly became a popular and widely used tool (Butcher, 2010; Neukrug, 2015). The MMPI was superseded in 1989 by the MMPI-2, and the Restructured Clinical scales were later added (Acheson & Thorpe, 2017; Forbey & Ben-Porath, 2007).

The MMPI-2-RF was published in 2008 and provides a streamlined alternative to the MMPI-2 (Ben-Porath & Tellegen, 2008). The MMPI-2 includes 567 true/false questions, while the newer MMPI-2-RF has only 338 true/false items and takes approximately half the amount of time to complete (35-50 minutes; (Acheson & Thorpe, 2017).

The MMPI-2-RF provides scores on 51 scales, including nine validity scales, three higher-order scales, nine restructured clinical scales, 23 specific problems scales (divided into four domains: somatic, internalising, externalising, and interpersonal), two interest scales, and five scales designed to measure personality psychopathology (Acheson & Thorpe, 2017).

California Personality Inventory (CPI)

The CPI was developed in 1957 by Gough to help clients achieve a better understanding of self through the use of true-to-life descriptions and clear, everyday language. All versions of the CPI are based on folk concepts of personality (Dean & Freeman, 2010).

“The CPI was the first well-developed measure of normal personality designed to predict high-level effectiveness in important areas of human performance (as contrasted with the factor analysis focus on measuring traits)” (R. Hogan & Roberts, 2001, pp. 7-8).

As such it was widely used in organisational contexts prior to the advent of contextualised occupational tools.

The latest version is the CPI 260. Released in 2008, the CPI 260 is based on and provides an abbreviated form of the third edition (published in 1995), with 260 forced choice items rather than 434. The CPI 260 takes approximately 25-30 minutes to complete, and provides scores on 29 scales. Twenty-six of the scales fall into five categories; the remaining three scales are higher order measures (Dean & Freeman, 2010).

16 Personality Factors Questionnaire (16PF)

The 16PF was originally developed in 1949 by Raymond Cattell. It was designed as a broad measure of personality, to predict a wide range of behaviours across a range of settings (including clinical, education and organisational; Carrington-Rotto & McLellan, 1995).

The 16PF is based on 16 traits identified by Cattell via factor analysis (Neukrug, 2015).

The latest version is the fifth edition, published in 1993. It has 185 forced-choice items and provides scores on 16 primary factors, 5 global factors and 3 validity scales (Carrington-Rotto & McLellan, 1995). The personality inventories 15FQ and 15FQ+, regularly used by New Zealand consultancies, measure Cattellian constructs and are derivative of the 16PF (see below).

Myers-Briggs Type Indicator (MBTI)

The MBTI was developed by Briggs and Myers in 1942, and has had numerous revisions since that time. It is based on Jung's 1923 theory of psychological types, which posits four dichotomous scales, leading to 16 possible personality types (Hess & Lanning, 2003).

The latest version, MBTI Step II Form Q, includes 144 forced-choice items and uses Item Response Theory scoring. It provides scores on the four dichotomies, as well as five subscales, or facets, for each dichotomy (Hess & Lanning, 2003).

NEO-PI-3 + NEO-FFI

The NEO was originally developed in 1978 by Costa and McCrae, to measure three domains of personality (Neuroticism, Extraversion, and Openness to Experience). Agreeableness and Conscientiousness were added in later revisions (Benson & Kluck, 2014).

In addition to measuring the Big Five, the NEO provides scores on 30 facets of personality (six per Big Five domain).

The latest revision (NEO-PI-3) was published in 2005 and includes 240 items using five-point Likert-type scales. It takes approximately 30-40 minutes to complete, and includes options for both self-report and observer ratings.

The NEO-FFI is a short-form version of the PI-3; it does not provide facet scores (Benson & Kluck, 2014).

Contextualised tools

Occupational Personality Questionnaire (OPQ)

For a detailed summary of the OPQ, including validation and reliability research, see Chapter 4: Sub-Project A: Method and Results.

Hogan Personality Inventory (HPI)

The HPI was developed by Hogan and Hogan in the 1980s and is based on socio-analytic theory. It provides a measure of normal personality and is designed for use in personnel selection and other organisational contexts (Axford & LoBello, 1998).

The HPI provides scores on seven primary scales, six occupational scales, a validity scale and a positive impression management index.

The second edition was published in 1995 and contains 206 true/false items (R. Hogan, Hogan, & Warrenfeltz, 2007). It takes approximately 15-20 minutes to complete.

15FQ+

The 15FQ was published in 1992 by Psytech International, and is based on 15 of the 16 personality dimensions posited by Cattell (see 16PF, above).

The current version, 15FQ+, added a sixteenth dimension, Intellectance. The 15FQ+ assesses behaviour across 16 personality traits and five higher order traits which are similar to the Big Five (Psytech International, 2016).

The 15FQ+ includes 200 items that use a three point rating scale (Yes/?/No) and takes approximately 30 minutes to complete.

2.2 Selection

As investigated in this thesis, personality assessment is one among many possible procedures of selection. Personnel selection is the real-life context within which job applicants have to face personality tests. The following sections address practices of test use that contrast with research evidence. Additional controversies are also highlighted.

2.2.1 Development of the selection process

Selection, as defined by Gatewood, Feild and Barrick, is the "process of collecting and evaluating information about an individual in order to extend an offer of employment" (2016, p. 3). From an applicant's perspective, selection is all that is experienced between applying for a job, and being offered that job or rejected. For an organisation, much of the work of selection takes place via development of the selection process well before the involvement of applicants (Roe, 2005).

The development of a best-practice selection process includes the following stages (Gatewood et al., 2016):

- Job analysis (including identification of work-related characteristics necessary for performance of the job)
- Selection, development and validation of assessment devices
- Use of assessment devices in the processing of applicants

Job analysis

Also referred to as work analysis[6], occupational analysis, and job specification, job analysis is typically defined as the systematic collection and analysis of work-related information regarding a job (Gatewood et al., 2016; Pearlman & Sanchez, 2010; Voskuijl, 2005).

This is likely to include tasks, responsibilities and work output; knowledge, skills, abilities and other personal characteristics; and, the wider organisational context (Cascio & Aguinis, 2011; Dachler, 1989; Pearlman & Sanchez, 2010).

The primary objective[7] of job analysis in the development of a selection process is to identify the characteristics required for successful job performance. Once identified, these become predictor variables that inform the choice and use of assessment devices, and the evaluation of applicants (Greuter & Algera, 1989; Pearlman & Sanchez, 2010; Voskuijl, 2005).

[6] Some texts refer to work analysis as a modernised version of job analysis that takes into account a wider context. The use of job analysis here is synonymous with the broader definition of work analysis as per Gatewood et al. (2016).

[7] Job analysis also provides useful information about the job to aid applicants, and can be used for the later assessment of job performance for the successful applicant (Gatewood et al., 2016; Jeanneret & Zedeck, 2010; Pearlman & Sanchez, 2010).

Selecting, developing and validating assessment devices

The assessment devices chosen for the collection and evaluation of application information must a) measure the predictor variables identified via the job analysis, and b) differentiate between applicants (Gatewood et al., 2016). “The whole enterprise of personnel selection is dependent on the existence and accurate measurement of individual differences among applicants.” (Vinchur & Koppes Bryan, 2012, p. 11).

Gatewood et al. (2016) provide a comprehensive list of factors to consider for the purpose of choosing selection procedures. When it comes to personality assessments, attention should be paid to selecting the test that is most appropriate for the job being selected for.

“Personality measures have consistently shown stronger predictive power in some settings and jobs than in others, and a careful analysis of where to use personality measures, which constructs to apply, and what criteria to choose makes all the difference. Unlike cognitive tests, it is not sensible to apply personality tests to all jobs and situations. Rather it is necessary to know a good deal about the job and the situation in which tests will be applied before a firm prediction about the relevance and usefulness of particular personality inventories can be made” (Murphy, Deckert, & Hunter, 2013, p. 645).

Once specific selection procedures have been identified, they are either sourced externally or custom-developed. Each selection procedure must then be validated within the current context of the hiring organisation, as the generalisability of research findings (or other evidence) in relation to selection procedures cannot be assumed (Kehoe & Murphy, 2010).

“It is really only after the validation phase has been completed that one has evidence that the information collected by the selection instrument is indicative of job performance, and therefore, useful for choosing among applicants” (Gatewood et al., 2016, p. 12).

Use of assessment devices in the processing of applicants

The final step in the development of a selection process is confirmation of how the information provided by the chosen selection procedures will be used.

An important decision concerns which parts of the information provided by selection procedures will be considered in the evaluation of applicant suitability.

For example, if the personality trait Conscientiousness is one of the predictor variables identified via the job analysis, and the OPQ-32R is the specific selection procedure chosen to

measure that variable, the organisation must decide whether to restrict the information provided by the OPQ-32R to only applicant scores on Conscientiousness, or not.

As quoted in the Introduction chapter, Carless (2009) recommends that full personality profiles are not shared with the selection panel, yet this appears to be common practice.

Another factor to be considered is how to weight the data elicited by each individual predictor variable. All information can be treated equally, and added together to form a score per applicant. Or, different pieces of information can be weighted asymmetrically, so that one predictor variable contributes more to the applicant's total score than another (J. W. Johnson & Oswald, 2010).

2.2.2 Usage of personality assessment results in selection

As noted earlier, there is widespread concern regarding the documented gap between research evidence and the application of personality assessments in selection (Klehe, 2004).

In contrast to the process just described, organisations (including in New Zealand) often fail to complete job analysis, fail to select personality assessments on the basis of identified predictors, fail to complete validation of the selected assessments, and fail to use the information supplied by those devices in an evidence-based manner (Taylor et al., 2002; Taylor et al., 1993).

Klehe (2004) details a number of reasons for the so-called scientist-practitioner gap. Possible explanations in specific relation to the use of personality assessments are suggested below.

Harris et al. (1996) found that New Zealand personnel consultants had a poor understanding of selection method validity and its relationship to selection, with 33% of respondents unsure of the meaning of validity.

This is of particular concern considering that test vendors often market their tools via reference to the validity of their tests. What may be misunderstood are the assumptions and/or validation efforts that must be made in order to generalise these results to a specific organisation's selection context.

The nature and requirements of validation have attracted theorists from different schools of psychometrics (Kaplan & Saccuzzo, 2018; Messick, 1990). Classically, validating a test involves gathering empirical evidence toward content-, criterion-, and construct-related validity.

Specifically, a predictor test has to be validated for criterion-related validity where a validity coefficient will enable estimation of the proportion of variance in job performance accounted for by this test.

Slaney (2017) proposes that, “validation is the process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use. It involves the accumulation of evidence from multiple sources in order to provide a sound scientific basis for a proposed interpretation of a test score for a specific use. It is incorrect to use the unqualified phrase “the validity of the test"” (Slaney, 2017, p. 275).

In addition, Rynes et al. (2002, p. 160) found that only 42% of 959 American Human Resources (HR) professionals surveyed, “correctly disagreed with the statement that there is very little difference among personality inventories with respect to how well they predict job performance”. A later replication study found similar results with only 52% of 626 Dutch HR professionals correctly disagreeing with the same statement (Sanders et al., 2008).

Findings like these contribute to apprehension that those selecting and administrating personality assessments in selection are not required to have a formal education in either psychology or psychometrics (Boylan, 2014; Taylor et al., 1993).

2.2.3 Effects of selection

The recommendations described above regarding best-practice selection stem from a large quantity of research into selection-specific outcomes for organisations, and promise to optimise for the organisation their selection of the most suitable applicant, while avoiding legal and ethical complications.

Anderson (2004) notes the disparity between a vast research base in relation to organisational perspectives on selection, and a relative dearth of studies taking the perspective of applicants.

“In contrast to this plethora of organization perspective research there has been a paucity of studies adopting an applicant perspective to investigate candidate reactions, applicant decision making, and the potential longer term psychological effects of exposure to selection methods” (Anderson, 2004, p. 2).

Applicant reactions research

Applicant reactions is a growing area of research attempting to redress this imbalance by investigating selection from the perspective of applicants.

There are multiple theoretical frameworks for understanding applicant reactions, such as justice expectations and attribution theory (Bell, Ryan, & Wiechmann, 2004; Ployhart & Harold, 2004). However, the majority of research has utilised the organisational justice framework pioneered by Gilliland (1993; McFarland, 2013; Ployhart & Ryan, 1997).

This framework proposes that situational and personal conditions (such as test type, HR policies, behaviour of HR personnel, and performance expectations) affect applicants' perceptions of whether procedural and distributive rules are satisfied or violated. The model then links these perceptions of fairness with applicant and organisational outcomes (such as job application decision, test motivation, self-esteem and self-efficacy, endorsement of the company's product, job satisfaction, performance, organisational citizenship behaviour, and organisational climate; Gilliland, 1993).

With regard to personality assessment for selection, the organisational justice framework accounts for negative applicant reactions to personality assessments via rule violations in relation to the following procedural rules: job relatedness, selection information, and propriety of questions (Rosse, Miller, & Stecher, 1994).

However, despite the focus on applicants, much of the research in this area has continued to emphasise the influence of applicant reactions on organisational outcomes, and has largely ignored the effects of selection processes on applicant wellbeing.

"Stated bluntly, if we were to be challenged tomorrow to demonstrate a lack of negative psychological effects upon applicants caused by exposure to various selection methods, there is considerable doubt that we could do so" (Anderson, 2004, p. 3).

While there have been studies within applicant reactions research regarding how elements within a selection process affect perceptions of fairness, which can then have an effect on self-perception (Gilliland, 1993), no studies have surfaced that have directly investigated whether personality assessment within a selection process affects applicants' self-perception.

Side-effects of personality assessment for selection

In addition to the more obvious selection-related outcomes for organisations and applicants discussed above, some authors have expressed concerns regarding a number of potential side effects or by-products of the use of personality assessment for selection. One of these,

reification of personality and personality traits has been noted above; three others are summarised below.

Evaluation of personality

The use of personality assessments as an evaluative tool for the purposes of employee selection implies that some personality traits or profiles are more valuable than others[8].

As discussed earlier in this chapter, the use of personality assessments within selection processes should be clearly linked to the results of job analysis, via which the personality traits or profiles sought after by the organisation have been identified as relevant to the specified job.

However, even when this is the case, the perceptions held by applicants may or may not correlate with the organisation's use of work-related characteristics in their appraisal of personality assessment results.

For applicants who are unsuccessful in their application for a job, via a selection process that included a personality assessment, there may be a perception that they have 'failed' the personality test, or have an otherwise undesirable personality.

Within this context, the proposal that there are 'no wrong answers' to personality tests (Gordon & Webb, 2014) may come across as disingenuous.

Generalised perceptions may also arise as inferences on the basis of the inclusion of the assessment itself.

Just as the assessment of personality can be argued to reify the existence of personality and personality traits, so too the use of assessment in selection as an evaluative tool may create and reinforce the belief that some personalities are held to be better than others.

Hanson (1993) argues that an intelligence-based, 'meritocracy' can produce devastating effects on self-esteem because those at the bottom become convinced they are there because they are inferior.

It is plausible that the evaluation and ranking of applicants within a selection process creates a meritocracy in relation to the job being applied for, and could therefore have a similar negative impact on self-esteem.

[8] This directly conflicts with Allport's original intention that personality would be a non-evaluative replacement for 'character' (Nicholson, 2003).

Homogeneity of employees

A related concern implicates personality assessments in the selection of overly homogeneous employees.

Whyte (cited in Hanson, 1993) claims that the use of placement tests following WWI produced conformity among executive employees by consistently rewarding extroversion, disinterest in the arts and cheerful acceptance of the status quo.

Quenk's admonition regarding inferences from personality data is revealing. "Interpreters of the MBTI assessment must also be wary of simplistic and incorrect uses of research data, particularly those showing type differences in such areas as career choice, managerial status, and leadership roles. A common error made (…) is to assume that the [personality] types who predominate in an endeavor are therefore more suited for it or 'better' at it. In fact, type theory predicts that individuals of different types will be differentially attracted to different occupations and work characteristics' (Quenk, 2009, p. 73). She later warns that personality theory underlying the test 'does not predict competence or satisfaction'.

A growing body of research points to the organisational benefits of diversity, including diversity of thought (Cairns & Preziosi, 2014; Deloitte Australia, 2011a, 2011b).

This may call for an expansion of job analyses and the selection of predictor variables beyond those traditionally linked with job performance.

"Once skills and qualifications are determined, most organizations spend time assessing which candidates best fit their organizational culture. In order to attract and hire for diversity of thought, organizations are going to need to expand their definition of cultural fit." (Cairns & Preziosi, 2014, p. 6).

Neuman, Wagner and Christiansen (1999) found that some specific personality traits contributed to team performance when they were homogeneous within a team, while other traits contributed more when they were heterogeneous.

Response construction

A cognitive proposal that sheds light on an additional risk in personality assessment has been termed the 'response construction' approach (Bogart, 1967; Morwitz, Johnson, & Schmittlein, 1993; Peterson, 2005; Wood, Conner, Sandberg, Godin, & Sheeran, 2014). Simply due to the experience of being questioned, individuals modify certain cognitions and may alter their behaviour.

For example, Morwitz, Johnson and Schmittlein (1993) found that survey respondents were more likely to purchase or not purchase items as a result of having been asked about them.

The mechanism of response construction (Feldman & Lynch, 1988) obeys principles of 'cognitive economy', whereby beliefs are not maintained in memory without an active purpose. Instead, they are newly (re-) constructed as prompted by communicative events, such as questioning.

It is assumed that elements of self-perception, including self-knowledge of personality, may not be represented prior to questioning such as items in a personality assessment that call for a response. Via response construction, assessment therefore has the potential to 'create' self-knowledge of personality.

"Belief, attitude, or intention can be created by measurement if the measured constructs do not already exist in long-term memory. The responses thus created can have directive effects on answers to other questions that follow in the survey. But even when counterparts to the beliefs, attitudes, and intentions measured already exist in memory, the structure of the survey researcher's questionnaire can affect observed correlations among them. The respondent may use retrieved answers to earlier survey questions as inputs to response generation to later questions (Feldman & Lynch, 1988, p. 421).

2.3 Self-perception

2.3.1 Theorising self-perception

The self

The concept of self is both complex and controversial (Baumeister, 1997; Robins, Tracy, & Trzesniewski, 2008; Wylie, 1974), with thousands of psychological articles published about it (Baumeister, 1998).

"Probably the term is rooted in such widespread common experience and basic linguistic, communicative needs that linguistic definitions may fail to do it justice. Even dictionaries are quite unhelpful for defining self. It is a word that everyone uses but no one defines" (Baumeister, 1998, p. 681).

Self is implicated in many theories of philosophy, psychology and psychotherapy, and it is central to contemporary Western culture and values (Hattie, 1992; Rose, 1996).

Rose and others detail the constructive evolution of current understandings of self and self-perception (Baumeister, 1997; Gergen, 2011).

"'The self,' whatever virtues of humanity and universality it may entail, thus appears a much more contingent, heterogeneous, culturally relative notion than it purports to be, dependent on a whole complex of other cultural beliefs, values and forms of life." (Rose, 1996, p. 6).

Defining self-perception

Despite the lack of clarity with regard to self, there are a multitude of psychological constructs predicated upon it.

Within this thesis, self-perception is used as an umbrella term to refer to constructs or concepts that are self-referent (Wylie, 1979) and based on self-beliefs or self-construals (Maehr, 2005).

Some of the many self-related constructs are as follows:

- Self-acceptance (MacInnes, 2006)
- Self-awareness (Silvia & Duval, 2001)
- Self-concept (Hattie, 1992; Marsh, 1990)
- Self-consciousness (Fenigstein, Scheier, & Buss, 1975)
- Self-efficacy (Bandura, 1977; Pajares & Schunk, 2005)
- Self-esteem (Rosenberg, 1986; Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995)
- Self-evaluation (Demo, 1992)
- Self-image (Kihlstrom & Klein, 1997)
- Self-insight (or self-objectification (Allport, 1955)
- Self-knowledge (Vazire & Wilson, 2012)
- Self-schemata (Markus, 1977)

Providing differential definitions for this group of constructs is problematic because they have been inconsistently used and defined throughout psychological theory and research (Hattie, 1992; Pajares & Schunk, 2005). In her review of self-concept research, Wylie referred to a lack of adequate conceptual and operational definitions (1979).

More recently, Marsh accused social scientists in this area of the jingle-jangle fallacy, whereby they use the same labels to refer to different constructs and different labels to refer to the same construct (Bembenutty, 2009).

Despite these complications, numerous investigations have demonstrated the importance of self-referent constructs for individual wellbeing and life outcomes (Ayub, 2010; Hau, Kong, & Marsh, 2003; Herbert, Manjula, & Philip, 2013; Marsh, Craven, & McInerney, 2005).

Self-concept and self-esteem

For this thesis, two constructs were selected to represent self-perception: self-concept and self-esteem[9].

A differentiation has often been made in which self-concept is construed as descriptive and self-esteem as evaluative, but Marsh argues that there is no evidence for this distinction. (Bembenutty, 2009)

Instead, for the purpose of clarity, it is assumed that self-concept is a multidimensional construct that represents self-perception in relation to specific content areas, and self-esteem is a representation of global self-perception, not linked to any domain (Marsh, 1990).

"For me, self-concept is represented as a multidimensional and hierarchical model of how individuals feel about themselves and it is heavily influenced by self-perceptions of competency, the evaluations of significant others, attributions for success and failure, and social comparisons. Self-esteem refers to how individuals feel about themselves without reference to any content area or domain, and it is at the apex of the self-concept hierarchy. In my multidimensional, hierarchical approach of self-concept, I treat self-esteem as the global component of self-concept—somewhat analogous to general ability or IQ that is at the apex of the ability hierarchy" (Bembenutty, 2009, p. 539)

Global versus domain-specific self-perception

[9] When specifically referencing the sub-projects within this thesis, self-perception will denote self-concept and self-esteem collectively.

Historically, self-perception has been approached from a global perspective, leading to negative reviews regarding the utility of self-concept in particular (Harter, 1996; Marsh, 1990).

More recently, significant progress has been made via multidimensional models that differentiate domains such as academic, social, emotional, and physical (Markus & Wurf, 1987; Marsh & Craven, 2006; Marsh & Shavelson, 1985).

For example, Marsh's (1990) multidimensional, hierarchical model of self-concept (based on Self-Description Questionnaire responses) places general self-concept at the top of a hierarchy; on the second level are Non-academic self-concept, Academic English self-concept and Academic Mathematics self-concept; and, on the final level are Physical Ability, Physical Appearance, Peer Relationships, Parent Relationships, Reading, General School, and Mathematics. (For an extensive review of theoretical models of multidimensional self-concept, see Marsh and Hattie (1996)).

Conversely, self-esteem is still meaningfully considered to be a global construct (Harter, 1996), though there are also studies exploring domain-specific self-esteem (Rosenberg et al., 1995).

Theoretical relationship between personality and self-perception

The conceptual entanglements that complicate clarification of both personality and self-referent constructs extend to the relationship between personality and self-perception.

Authors vary with regard to how each construct is linked to the other, with a clear dependency on how each construct has been defined.

Self-concept as part of personality

A number of authors present self-perception as a part of personality (Barenbaum & Winter, 2008; Grubbs, Wilt, Stauner, Exline, & Pargament, 2016; Judge & Bono, 2001).

McAdams and Pals (2006, p. 204) define personality as "(a) an individual's unique variation on the general evolutionary design for human nature, expressed as a developing pattern of (b) dispositional traits, (c) characteristic adaptations, and (d) self-defining life narratives, complexly and differentially situated (e) in culture and social context".

McCrae and Costa (2008) also include self-perception as a sub-component of 'characteristic adaptations' - a component of personality alongside 'basic tendencies' (which represent the Big Five personality traits).

Marsh et al. (2006) found that specific components of self-concept predicted substantial variance in personality factors.

"In the extreme version of the core-surface distinction, (core) personality factors are expected to cause (surface) self-concept factors, whereas self-concept is expected to have no causal effect on personality traits. In contrast to this extreme position, we predict the pattern of relations is likely to be reciprocal, such that personality factors and self-concept factors are each causes and effects of each other—although we concede that the causal effects of personality are likely to be stronger than those of self-concept on personality" (Marsh et al., 2006, p. 447).

Research has also demonstrated support for self-concept as a mediator of the influence of childhood maltreatment on adult personality pathology (Cohen, Leibu, Tanis, Ardalan, & Galynker, 2016).

Self-concept of personality

A second link between the two constructs is 'self-concept of personality' (or personality self-concept): a component of self-concept containing knowledge about one's own personality (Perugini & Banse, 2007).

Asendorpf, Banse and Mucke (2002) assert that this is what test-takers access when completing self-report personality inventories.

Previous discussions have explored theories in which self-report personality inventories access 'true' personality versus a self-presentation of personality. Self-concept of personality adds a third option - one that may or may not align with so-called true personality (Back, Schmukle, & Egloff, 2009; De Cuyper et al., 2017).

A further (structural) component of self-concept, self-concept clarity, is correlated with four of the Big Five personality traits (J. D. Campbell et al., 1996).

2.3.2 Assessing self-perception

One of the key critiques of the self-perception literature, both in Wylie's original and subsequent reviews, and that of Shavelson et al. related to the lack of methodological rigour of instruments for the assessment of self-perception (Shavelson, Hubner, & Stanton, 1976; Wylie, 1961, 1979).

Since that time, dedicated focus has resulted in significant improvements, with continued effort recommended (Hattie & Marsh, 1996; Keith & Bracken, 1996; Marsh, Craven, & McInerney, 2003).

Specific tools

The following sub-sections provide a brief description of a selection of self-concept and self-esteem assessment tools that include or target adults. All of these tools are self-report inventories.

Self-Description Questionnaire III (SDQ-III)

The SDQ-III is one of three tests in the SDQ series (I is designed for ages 5-12, II for ages 13-17 and III for ages 16+). The SDQ-I was developed in 1987 by Marsh, as a measure of pre-adolescent self-concept based on the Shavelson et al. model (Marsh & O'Neill, 1984).

The SDQ-III contains 148 items, in two sections. Section 1 has 136 statements that are responded to on an eight-point scale, while Section 2 has 12 items with nine-point response scales.

The test takes approximately 15-25 minutes to complete, and provides scores on 14 dimensions of self-concept, grouped into Academic, Non-academic and Global categories (Atlas, Gable, & Isonio, 1998).

Self-Perception of Adults (SPA)

The SPA was developed from the original Self-Perceptions Inventory by Soares, and is part of the SPI series. It is designed to measure an adult's concept of self and includes two perspectives: Self as a Person and Self as a Working Adult.

Each section has two options for item format, either 40 dichotomous trait pairs or 30 sentences. Response options are forced choice between 'very aligned' or 'more aligned' for one of the two traits (Crumpton & Farmer, 2010).

The SPA takes five-20 minutes to complete and provides scores on eight scales (four per perspective).

The Six-Factor Self-Concept Scale (SFSCS)

For a critical summary of the SFSCS, including validation and reliability research, see Chapter 4: Sub-Project A: Method and Results.

Tennessee Self-Concept Scale (TSCS)

The TSCS was first published in 1965 by Fitts, and was one of the first tools to measure multiple dimensions of self-concept (R. Brown & Hattie, 1998).

The current version is the second edition, released in 1996. It has three forms: Adult (82 items), Child (76 items) and a short-form that includes the first 20 items of either the Adult or Child form.

Test items use five-point scales, and each form takes 10-20 minutes to complete.

The TSCS:2 (excluding the short-form) provides scores on six substantive scales, which can be grouped into a Total score and a Conflict score. There are also three supplementary scales, and four validity scales (R. Brown & Hattie, 1998).

Coopersmith Self-Esteem Inventory (CSEI)

The Coopersmith Self-Esteem Inventory was published in 1967 by Coopersmith (1967). It was designed to measure evaluative attitudes towards the self across a range of settings including social, academic, family and personal (Sewell, Peterson, & Austin, 1985). There are currently three forms: School, School Short Form and Adult.

The Adult form was adapted by Ryden (1978) via minor modifications to make the language more appropriate to adults. It contains 15 items that use a two-point scale consisting of 'like me' and 'unlike me, and takes approximately 10 minutes to complete.

Rosenberg Self-Esteem Scale (RSES)

For a critical summary of the RSES, including validation and reliability research, see Chapter 4: Sub-Project A: Method and Results.

Self-Esteem Assessment (SEA)

The SEA was published in 2011 by PsychTests AIM (Buhs & Scott, 2014). It is designed to evaluate general self-esteem for ages 17 to adult.

The SEA includes 79 items with a five-point Likert-scale response format and takes approximately 30 minutes to complete. It provides scores on nine scales, including an overall score (Buhs & Scott, 2014).

2.3.3 Change in self-perception

Stability of self-perception

There is a significant body of evidence in support of the stability of self-perception. Research demonstrates that people will actively seek information that confirms their self-perception, and reject information that conflicts with it (Greenwald, 1980; Swann & Ely, 1984; Wylie, 1979).

"Swann (1985) and Swann and Hill (1982) found that individuals will go to great lengths to confirm their self-perceptions by attending most closely to information that fits their view of the self and by trying to arrange their environment so as to acquire further self-confirming evidence. Individuals also tend to reject or ignore those accounts of their behavior that differ from their own" (Hattie, 1992, p. 52).

Conversely, many studies have also demonstrated that situational factors can cause temporary changes to self-perception (Heatherton & Polivy, 1991).

In their discussion of self-concept as a sub-component of personality, McCrae and Costa (2008) assert that (unlike personality), self-perceptions are subject to change over the lifespan.

"Self-concept is a structural product of reflexive activity, but it is also susceptible to change as the individual encounters new roles, situations, and life transitions. The data reviewed in this paper suggest that: (i) self-evaluation generally becomes more favorable through the life-span; (ii) self-evaluation is represented by a "moving base-line" from which situational fluctuations emerge; (iii) self-concept is characterised by both stability and change over the life course; and (iv) environmental stability plays an important role in self-concept stability" (Demo, 1992, p. 303).

Antecedents of change to self-perception

There is documented evidence for a broad range of antecedents to change in self-perception, from experiencing a stroke to falling in love (Aron, Paris, & Aron, 1995; Ellis-Hill & Horn, 2000).

The follow sub-sections summarise the literature regarding two antecedents of change to self-perception that have potential for comparison with personality assessment for selection.

Positive and negative events

Stake, Huff and Zand (1995) conducted two studies in which students were exposed to guided imagery scenarios presenting either positive or negative self-relevant information, such as social acceptance or unexpectedly scoring poorly on an important test. Stake is the author of the SFSCS, which was used as the measure of self-concept in Sub-projects A and B for this thesis.

After each scenario, participants provided a rating to indicate the extent to which they anticipated that the scenario would enhance or lessen their view of themselves, as well as ratings to indicate the extent to which they anticipated that the scenario would affect their view of themselves in specific relation to each of the six components of self-concept measured by the SFSCS. Participants also completed the RSES measure of self-esteem (also used in Sub-projects A and B as the measure of self-esteem), prior to the experiment.

The purpose was to explore the apparent paradox between momentary changes in self-perception and the stability of trait self-esteem, by examining the relationship between them (Stake et al., 1995).

The investigation referenced self-schema theory to predict that shifts in self-perception would occur in alignment with pre-existing self-esteem. That is, so-called negative events would have a greater negative impact on the self-perceptions of participants with already low self-esteem.

The authors confirmed their hypotheses, finding that anticipated self-perception was impacted as expected by the positive and negative events, with greater shifts in self-perception occurring as a function of pre-existing self-esteem. Participants with low self-esteem were more likely to anticipate greater reductions in self-concept (and greater negative generalisation across components of self-concept) as a result of the negative events.

Induced behaviour and biased scanning

A range of social psychology experiments have demonstrated that manipulating subjects' behaviour (or self-presentations) can result in a change to self-perception in the direction of the manipulated behaviour/presentation (see Fazio et al., 1981; E. E. Jones et al., 1981; Markus & Kunda, 1986; Rhodewalt & Agustsdottir, 1986; Schlenker, Dlugolecki, & Doherty, 1994; Tice, 1992).

This body of research draws upon models and theories including cognitive dissonance theory (Festinger, 1957), self-perception theory (Bem, 1972), expectancy confirmation or self-fulfilling prophecy (Darley & Fazio, 1980), self-verification theory and behavioural confirmation (Swann & Ely, 1984).

For example, participants who were induced to portray themselves as emotionally stable (or emotionally responsive) showed a subsequent change in self-reported self-concept to be more emotionally stable (Tice, 1992).

Biased scanning, a variant of self-perception theory, is one of the posited mechanisms for the above changes to self-perception (E. E. Jones et al., 1981; Tice, 1992). This theory can also account for effects of interpersonal feedback on self-perception (Gecas, Calonico, & Thomas, 1974; Phylactou, 2000).

3 Hypotheses

This chapter presents the operationalisation of the dependent variable, followed by descriptions of the original research proposal and the amended research designs used in Sub-projects A and B. It concludes with the resulting hypotheses that have guided each sub-project within the current research.

3.1 Operationalisation of the dependent variable

As discussed in the Introduction chapter, the objective of this thesis was to explore the broad hypothesis that completing a personality assessment and/or receiving results/feedback from that assessment could cause a change to self-perception.

Two constructs were selected to represent self-perception: self-concept and self-esteem. As discussed in the previous chapter, the literature on self-referent constructs is complicated by inconsistent usage and definitions. These constructs were selected in accordance with Marsh's classification of self-concept as a multidimensional construct representing self-perception in relation to specific content areas, and self-esteem as a representation of global self-perception, not linked to any domain (Bembenutty, 2009).

The instruments chosen to assess these constructs were the Six-Factor Self-Concept Scale (SFSCS) and the Rosenberg Self-Esteem Scale (RSES). The SFSCS is one of the few self-concept assessment tools developed for use with adults; it rates six components of self-concept: Likeability, Task Accomplishment, Power, Vulnerability, Moral and Gifted, which can also be summed to provide a Total Self-concept score.

The RSES is a single factor assessment of global self-esteem. (For further discussion of each of these instruments, see the following chapter: Method and Results: Sub-project A).

The dependent variable for this project (change in self-perception) was represented by comparisons of before and after scores on both the SFSCS and RSES.

3.2 Changes to the research design

3.2.1 Original design

The original design for this project envisioned a longitudinal, within-subjects, observational study with three waves of measurement, using a large sample of job applicants recruited via

multiple participating organisations. It was hoped that the participant sample would span organisations using different personality assessments in their selection processes, so that the effects on self-perception of different assessment tools could be compared, and results generalised to a wider population.

The intention was that participating organisations already conducting personality assessment within selection processes would advertise the research project to their job applicants (as an adjunct to their standard communications).

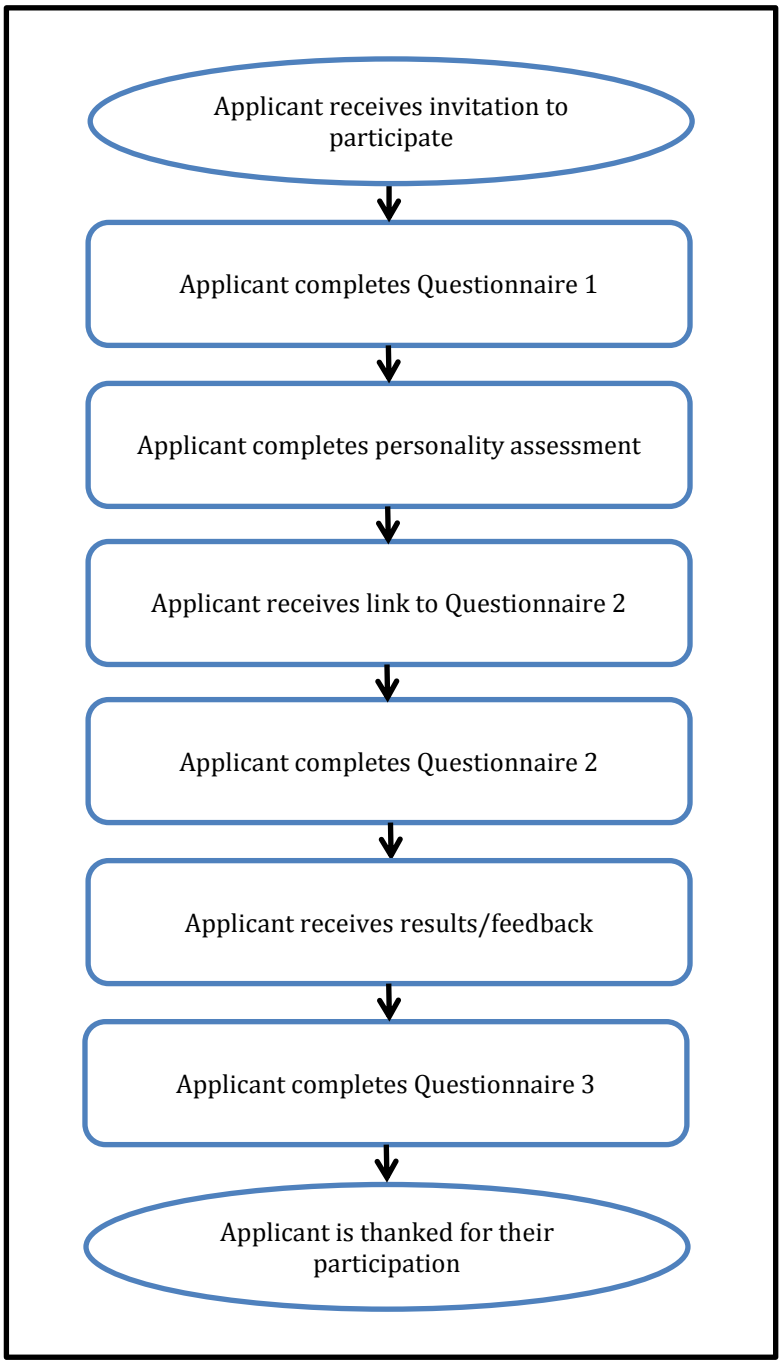
Self-selected participants would complete an online questionnaire prior to completing the personality assessment for the hiring organisation; then a second questionnaire following completion of the personality assessment; and a third questionnaire three months later.

By the time of the third and final wave of measurement it was hoped that at least some applicants would have received results/feedback from their personality assessment and therefore for these participants the final questionnaire would elicit an assessment of self-perception following receipt of results/feedback.

It was expected that some participants would show observable change in self-perception following completion of the personality assessment, but that for greater numbers of participants (and at greater magnitudes), change would occur after they had received results/feedback from the personality assessment.

The intention was to conduct quantitative analyses of the scores on the SFSCS and RSES, such as dependent-measures t-tests and MANOVAs.

Figure 1: Original design sequence



For the reasons outlined above, the original design was preferable to the eventual design of both Sub-projects A and B. It is recognised that the reduced designs restrict opportunities for analysis. These issues are reviewed in greater detail in the Discussion chapter.

3.2.2 Sub-project A

In recruiting for the original design, only one organisation was eventually able to participate and provide participants.

As a result of this, the participant sample for Sub-project A was comprised of job applicants applying for jobs at one organisation only, and all participants completed the same personality assessment as part of the selection process. Due to this, comparison of the effects of different personality assessments was not possible.

Early on in data collection the project was also hampered by low participant commitment. Participants attempted the first questionnaire, with reduced numbers completing the second, and at that point participation halted, with only one participant ever commencing the final questionnaire.

This outcome forced a reconsideration of the design, resulting in the final inclusion of only two waves of measurement: Before and after completion of the personality assessment.

Despite an intuition that results/feedback were likely to have a higher impact on self-perception than the assessment alone, it was nevertheless deemed important to capture the potential effects of the assessment itself due to the low frequency of applicants in real life who receive results/feedback following completion of a personality assessment as part of a selection process.

Also, by this time the project had been approved by the participating organisation and set in motion. Altering the project design to remove the final questionnaire was a simple solution that provided neither delay nor inconvenience to the participating organisation.

Attempts were made to initiate an auxiliary study, with another two waves of measurement occurring before and after personality assessment results/feedback were received. Ultimately, no organisations were able to participate and the auxiliary study did not go forward.

3.2.3 Sub-project B

Later, due to low overall participant numbers in Sub-project A despite a lengthy data collection period, organisational sponsorship for a second investigation was sought and gained from a prominent personality assessment vendor.

The design for this sub-project was very similar to the first, with only two major differences as follows. First, at the request of the sponsor, the participant sample was comprised of

university students only. Second, the two waves of measurement occurred before completion of a personality assessment (as in Sub-project A), and following both completion of the assessment and receipt of a written results/feedback summary (different to Sub-project A).

The constructs under investigation (self-concept and self-esteem) remained the same as for Sub-project A, as did the assessment tools used. By extension, the intended quantitative analyses were also the same[10]. Even the questionnaires were almost identical, altered only to remove references to the real-world selection process in which participants in Sub-project A had been involved. By chance, the personality assessment provided by the sponsor in Sub-project B happened to be the same personality assessment tool completed by participants in Sub-project A.

3.3 Hypotheses

The following sections list the hypotheses that were tested[11] in Sub-projects A and B.

Anecdotal evidence regarding the effects of personality assessment for selection suggests that for some the experience is positive, and for others negative. For these reasons, the major hypotheses were non-directional. Without a research-base in this area to build upon, the investigations in these sub-projects were exploratory in nature; therefore multiple minor hypotheses were proposed. These have been grouped together under Hypothesis 1a and Hypothesis 1b, as below.

3.3.1 Sub-project A

Hypothesis 1

Completion of a personality assessment as part of a selection process will be associated with a change in:

[10] During the course of Sub-project B, the author came across a number of critiques of traditional statistical analyses used in psychological research. These sources inspired an attempt to conduct additional, non-quantitative analyses such as pattern analysis of the results. Ultimately, however, the data collected was not well suited to this aim and there was insufficient time to explore this alternative approach.

[11] Hypotheses in relation to the aborted third wave of measurement in Sub-project A, and to qualitative analyses (see Footnote 10) have been omitted.

- Any or all of the following components of self-concept as measured by the Six-Factor Self-Concept Scale: Likeability, Task Accomplishment, Power, Vulnerability, Gifted, Moral; and/or,
- Total self-concept as measured by the Six-Factor Self-Concept Scale; and/or,
- self-esteem as measured by the Rosenberg Self-Esteem Scale

Hypothesis 1a

Positive or negative change to self-concept and/or components of self-concept and/or self-esteem will be influenced by any of the following variables:

- No. of roles applied for in last month
- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results
- Confidence in correct interpretation of ability assessment results
- Sex
- Age
- Ethnicity
- Current employment status
- Reason for applying for role
- Expected salary
- Previous experience of personality assessment(s)
- Reason(s) for prior completion of personality assessment(s)

3.3.2 Sub-project B

Hypothesis 2

Completion of a personality assessment and receipt of results/feedback will be associated with a change in:

- Any or all of the following components of self-concept as measured by the Six-Factor Self-Concept Scale: Likeability, Task Accomplishment, Power, Vulnerability, Gifted, Moral; and/or,
- Total self-concept as measured by the Six-Factor Self-Concept Scale; and/or,
- self-esteem as measured by the Rosenberg Self-Esteem Scale

Hypothesis 2a

Positive or negative change to self-concept and/or components of self-concept and/or self-esteem will be influenced by any of the following variables:

- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results
- Sex
- Age
- Ethnicity
- Previous experience of personality assessment(s)
- Reason(s) for prior completion of personality assessment(s)

4 Sub-project A

4.1 Sub-project A: Method

4.1.1 Participants

Participant source

Participants were New Zealand job candidates who were shortlisted by the participating hiring organisation (a small-medium government department located in Wellington, New Zealand) during the period April 2014 – October 2015, and who received and responded to a research invitation provided via email from the participating hiring organisation.

Method of recruitment

The invitation was conveyed via email as an addition to an existing email template that notified candidates of a requirement to complete psychometric testing (after being shortlisted for a role for which they had applied).

Participant characteristics

See Tables 1 - 5 for distributions of participants in relation to sex, age group, ethnicity, and previous experience (including reasons for completion) of personality assessments.

Table 1: Sex – Sub-project A

Sex	Count	Percentage
Female	18	66.67
Male	9	33.33

Table 2: Age group – Sub-project A

Age group	Count	Percentage
Under 16	0	0.00
16 to 24 years	3	11.11
25 to 34 years	17	62.96
35 to 44 years	2	7.41

Age group	Count	Percentage
45 to 54 years	3	11.11
55 to 64 years	2	7.41
65 years and over	0	0.00

Table 3: Ethnicity – Sub-project A

Statistics New Zealand Level 1 classification[12]	Count	Percentage
European	22	81.48
Maori	2	7.41
Pacific Peoples	0	0.00
Asian	3	11.11
Middle Eastern	0	0.00
Latin American	0	0.00
African	0	0.00

Table 4: Previous experience of personality assessment(s) –Sub-project A

Previous experience of personality assessment(s)	Count	Percentage
Yes	17	62.96
No	10	37.04

Table 5: Reason(s) for prior completion of personality assessment(s) – Sub-project A

Reason	Count	Percentage
As part of a job application	11	64.71
Personal development (at work)	10	58.82
Personal development (outside work)	6	35.29
In school/university setting	5	29.41

[12] Ministry of Health (2004)

The most commonly completed personality assessment (of the 14 specifically queried) was the Myers-Briggs Type Indicator (MBTI; version not specified) with a frequency of 23 completions across 13 participants. The second most frequently completed assessment was the Occupational Personality Questionnaire (OPQ; version not specified), with six completions across five participants.

Exclusionary criteria and number of participants excluded

Participants were excluded if they completed the personality assessment before completing Questionnaire 1 (as this precluded the possibility of measurement of self-perception prior to the assessment). Participants were also excluded if they did not complete both Questionnaire 1 and Questionnaire 2.

Of the 73 candidates who accepted the invitation to participate, 14 were excluded because they completed the personality assessment prior to completing Questionnaire 1, and a further 13 were excluded for not completing the first questionnaire. Of the 46 remaining candidates who were sent the link to Questionnaire 2, 19 were excluded for non-completion.

Whether participant group was representative

Sub-project A did not include a process of sampling from a population, let alone representative sampling (Field, 2009; Frankfort-Nachmias, 1996; Hansen, Hurwitz, & Madow, 1993).

Representativeness is examined in detail in the Discussion chapter.

Acceptance rates for participants solicited

Because the invitation to individual participants was sent by the participating hiring organisation rather than the researcher it was not possible to determine the exact proportion of applicants who accepted the invitation (to participate) out of the total number of those who received the invitation.

How participants were assigned to conditions

There was only one condition in this study, with all participants going through the same process. The study was an observational study; the researcher did not assign participants to

the condition, rather participants self-selected to participate or not participate. Participants who completed both questionnaires were entered into a draw with five chances to win a $100 'Prezzy Card'.

Participating organisations were offered a results summary to be provided at the conclusion of the research.

4.1.2 Data collection

Contacting organisations

Contacts at 37 organisations were approached with a three-question, email-based survey asking about their organisation's use of personality assessments in selection. The purpose of this preliminary research was to discover the scope of personality assessment use in selection across a number of New Zealand organisations, and in so doing to identify prospective organisations for participation in Sub-project A.

The organisations that were approached represented a 'sample of convenience'; they were organisations where a representative known to the researcher (or the researcher's immediate contacts) worked in HR, or was able to communicate directly with HR.

The questions within the survey were as follows:

"Does your organisation use personality tests for selection (administered either by its own HR division, or via consultancy firms or contractors)?

Could you please name the test(s) used?

For which job(s) or kind(s) of jobs do you select applicants by means of a personality test, and what is the approximate number of applicants tested annually (e.g. 2012)?"

Of the 37 organisations contacted, 25 provided a response, and of those, 19 organisations affirmed use of personality assessments in selection for some or all vacancies. A summary of the results of this survey is shown in Appendix 2.

The organisations that disclosed use of personality assessments in selection (as well as three consulting agencies known to conduct personality assessments on behalf of organisations) were then contacted to request participation in Sub-project A.

While several organisations expressed interest in participation, only two organisations were eventually able to do so. These organisations commenced passing on to shortlisted job candidates the invitation to participate in the research, as described above.

All three consulting agencies also agreed to participate by advertising the research with selected client organisations, however no organisations approached by the consultancies agreed to participate.

Ultimately, all participants came from only one organisation – the second participating organisation attributed this to lack of interest from their job applicants, rather than any technical or other issue.

Design

The study employed a repeated-measures design, comparing measures of individuals' self-perception before and after completion of a personality assessment. As detailed in the Hypotheses chapter, the original design of the study included additional waves of measurement, but had to be curtailed due to low participant commitment.

Repeated measures refers to the use of one participant group, all of whom experience both independent variable conditions (Alexander & McGarty, 2014).

The repeated measures factor had two levels:

- Before completing a personality assessment for selection
- After (i.e. within a few days of) completing a personality assessment for selection

This research design can be susceptible to order effects (also called carry-over effects), including practice effects, fatigue effects, history effects and maturation (Nestor & Schutt, 2015; Wilson & Joye, 2017). The potential impacts of these are discussed in the Discussion chapter.

Sequence

The sequence for participants was as follows:

1. Each participant receives an email from the participating hiring organisation notifying him or her that they have been shortlisted for a job they have applied for and that they are required to complete psychometric testing as part of their application. The link to the online personality assessment is included in the email.
   The email also contains an invitation to participate in the research project for five chances to win $100. The link to Questionnaire 1 is included in the invitation, along with clear instruction to complete Questionnaire 1 prior to completing the personality assessment.

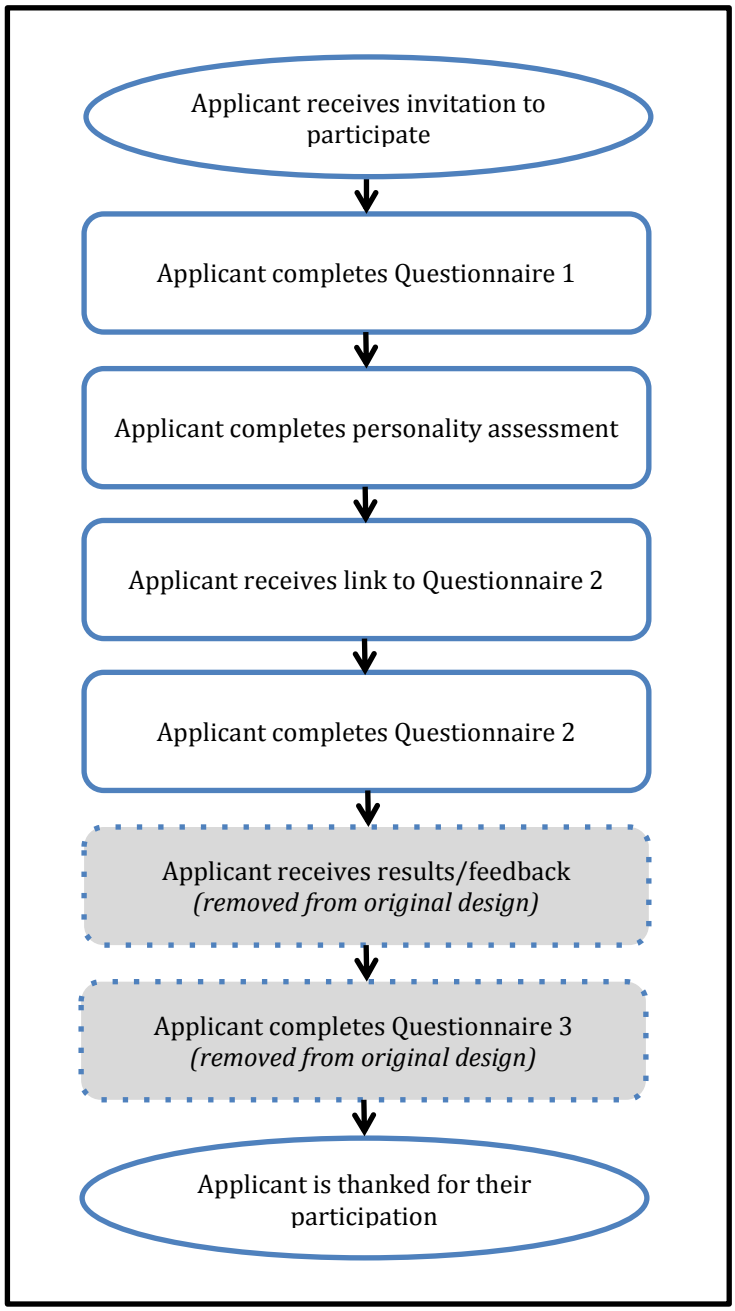
2. The participant completes Questionnaire 1.

3. Following completion of Questionnaire 1, each participant receives an email from the researcher to thank him or her for their participation and to request that they complete Questionnaire 2 as soon as possible following the personality assessment. The link to Questionnaire 2 is included in this email.

4. The participant completes Questionnaire 2.

Instructions and order of events were the same for all participants. See Figure 2.

Figure 2: Sub-project A sequence



Instruments

Instruments for data collection used in this sub-project were two self-report 'questionnaires' administered via the online tool Qualtrics Surveys. A more accurate description is 'merged data collection instrument', however for ease of communication, these were labelled as questionnaires for participating organisations and individual participants.

Participant responses and associated metadata were downloaded from Qualtrics Surveys and imported into IBM SPSS.

Each ‘questionnaire’ comprised a range of forced-choice (categorical, nominal and ordinal level) items and free-form items used and/or created by the researcher, as well as full scales of the Six-Factor Self-Concept Scale and the Rosenberg Self-Esteem Scale.

Questionnaires and questionnaire items were administered in the same order for all participants. The variables measured within the questionnaires were as follows.

- Demographics – sex, age group, ethnicity
- Context of job application – “current employment status”, “reason for applying for role”, “expected salary”
- Expectations regarding psychometric assessments within the selection process – “expectation of personality assessment as positive/negative experience”, “confidence in correct interpretation of personality assessment results”, “confidence in correct interpretation of ability assessment results”
- Previous experience of personality assessment(s) – “previous experience [Y/N]”, “reason(s) for prior completion of personality assessment(s)”, “specific tests completed”, “number of times completed”
- Self-concept (pre-assessment)
- Self-esteem (pre-assessment)
- Experience of completing personality assessment – “experience of personality assessment as positive/negative”, “difficulty completing personality assessment”, “confidence that personality assessment results were interpreted correctly”, “personality assessment allowed true and complete picture”, “questions that did not apply”, “consistency of responses”, “relevance to job”
- Experience of receiving results/feedback (not completed by any participants)
- Impact of completing personality assessment – “considered something new”, “thought differently about self”, “would answer differently in future”, “personality assessment of personal use”, “changed motivation”, “changed optimism”
- Outcome of job application – “successful/unsuccessful”, “accepted/did not accept offer”
- Self-concept (post-assessment)
- Self-esteem (post-assessment)

Numeric values were assigned to response options for (non-freeform) researcher-created items within the questionnaires, according to the assumed scale of measurement for each

item. Assignment of numeric values to response options for the Six-Factor Self-Concept Scale and Rosenberg Self-Esteem Scale is detailed in the sub-sections below.

Six-Factor Self-Concept Scale

Description

The Six-Factor Self-Concept Scale (SFSCS; Stake, 1994) is a multi-dimensional measure of self-concept for adults. The scale has 36 items, for each of which respondents are asked to rate the extent to which a descriptive phrase (e.g. "Fun to be with") is true of them, by selecting one of seven response options. The response options (and associated scores in brackets) are as follows: 'never or almost never true of me' (1), 'usually not true of me' (2), 'sometimes but infrequently true of me' (3), 'occasionally true of me' (4), 'often true of me' (5), 'usually true of me' (6), and 'always or almost always true of me' (7) (Stake, 1994).

Each of the 36 items belongs to one of six subscales: Likeability (six items), Task Accomplishment (six items), Power (seven items), Vulnerability (six items), Morality (six items) and Giftedness (five items). The items relating to each subscale are distributed throughout the scale in a non-uniform order. Scores for each subscale are achieved by adding together the scores for the items of that subscale. For the purpose of comparison of subscale scores, the Power and Giftedness scales are weighted (by .857 and 1.2 respectively).

A total self-concept score is achieved by adding scores for all subscales except for Vulnerability, which is subtracted from the summed total of the other subscales.

This method of scoring the SFSCS implies that the construct (self-concept) is assumed to be quantitative, and that item responses represent interval-level data.

What the SFSCS measures

Stake defines self-concept as, "the domain of self-descriptions that have a self-evaluative connotation" (Stake, 1994, p. 56). The SFSCS measures multiple dimensions within these self-evaluations, and at a mid-level of specificity (i.e., it is more specific than global self-concept, but less specific than self-concept within a particular domain, such as academic self-concept). This approach maximises generalisability and sensitivity (Stake & Eisele, 2010).

In keeping with the pursuit of maximum generalisability, the domains of self-concept measured by the SFSCS are intended to be those with universal relevance to all adults, and

within all settings (Smalley & Stake, 1992). The two overarching themes within the scale are harmonious interpersonal functioning (e.g. the ability to engender liking and appreciation from others, moral goodness and virtue), and agentic functioning (e.g. aptitude, competence, mastery, and coping ability); each of the subscales relates to one of these themes.

Likeability and Morality pertain to interpersonal functioning. Likeability refers to the ability to develop positive relationships with others, while Morality refers to universal qualities of goodness and virtue. Task Accomplishment, Power, Giftedness and Vulnerability (reversed scoring) pertain to agentic functioning. Task Accomplishment refers to efficient and capable management of tasks; Power refers to strength, leadership and persuasiveness; Giftedness refers to innate rather than learned abilities or achievements[13]; and, Vulnerability refers to tendencies to feel criticised and criticise the self, and to find it difficult to perform under pressure (Smalley & Stake, 1992; Stake, 1994).

Each subscale is a self-evaluative dimension with a positive and negative pole; Stake and Eisele (2010) affirm that this differentiates them from personality traits, which are merely descriptive.

Test development

The SFSCS was published in 1994 after extensive testing (Stake, 1992, 1994). One hundred and fifteen items relating to the themes described above were given to four samples of (476) undergraduate students. The original items related to the Likeability subscale came from the Social Self-Esteem Scale, also created by Stake (Stake, 1985, 1994). Factors were identified using exploratory factor analysis, with Cattell's scree test used to revise the items down to a final set of 36 items. The final extraction yielded the six dimensions or subscales detailed above.

These six subscales were tested on a sample of 365 adults of different ages and occupational and socioeconomic categories. They were confirmed by both exploratory factor analysis and confirmatory factor analysis (Stake, 1994).

Predicted relationships in order of strength were demonstrated between self-concept as measured by the SFSCS and measures of social desirability, wellbeing, and self-esteem (Stake,

[13] Note that the Gifted subscale of the SFSCS does not relate to the field of giftedness within psychology, e.g. See Renzulli (1978).

1994). Results were also correlated with self-ratings of childhood memories and perceptions of recent behaviour or events, and with observer ratings of the SFSCS.

The SFSCS has test-retest reliability of .97 for total scores, and between .74 and .88 for subscales (Stake, 1994).

Psychometric properties

There are a number of studies that have employed the SFSCS; where indications of reliability and validity have been reported, these are included below.

Reliability

Ayub (2010) investigated the relationship between self-concept and life-satisfaction among adolescents, using the SFSCS as the measure of self-concept. Her study demonstrated internal consistency in the SFSCS of .887.

Zimmerman, Ownsworth, O'Donovan, Roberts and Gullo (2017) found internal consistency of the subscales of the SFSCS ranging from good to excellent (.76 - .94) in their sample of 42 adults with autism spectrum disorder.

Vecina, Chaćon and Pérez-Viejo (2016) found internal consistency of .824 and .812 for the Moral subscale of the SFSCS in their samples of violent and non-violent men.

Herbert, Manjula and Philip (2013) studied correlates of resilience including self-concept in adult children of parents with schizophrenia. They found strong internal reliability alpha coefficients for all subscales of the SFSCS, ranging from .76 to .86.

Lu and Chang (2011) utilised the Moral subscale of the SFSCS and found high internal consistency (.85) for the subscale.

Validity

Herbert, Manjula and Philip (2013), referenced above, reported significant ($p < .01$) positive correlations between resilience and Likeability, Morality, Task Accomplishment and Giftedness (as well as Power ($p < .05$)), and a significant negative correlation between resilience and Vulnerability. Stepwise regression analysis demonstrated that all subscales (with the exception of Power) predicted resilience ($p < .01$).

Two studies using the same participants (Van Wyk, Boschoff, & Cilliers, 2003; Van Wyk, Boshoff, & Bester, 2003), found correlations between self-concept measured by the SFSCS, and entrepreneurial attitudes and job involvement. However, Principle Factor Analysis did not support the six factors but rather three factors, labeled by the authors Power (14 items), Task Accomplishment (12 items) and Likeability (6 items), with four of the original 36 items removed.

Conversely, Yanico and Lu (2000) conducted validity testing on the SFSCS with a sample of racial/ethnic minority college women and found support for the construct validity of the measure, as well as replication of the factor structure, and convergent validity with other measures.

The full scale of the SFSCS can be retrieved from PsycTESTS: doi: 10.1037/t07156-000.

Rosenberg Self-Esteem Scale

Description

The Rosenberg Self-Esteem Scale (RSES) is a widely-used measure of global self-esteem (Rosenberg, 1965). The instrument has 10 items, for each of which respondents are asked to rate their level of agreement with a statement by selecting one of four response options. Five of the items are worded positively, e.g. "I feel that I have a number of good qualities" and five negatively, e.g. "All in all, I am inclined to feel that I am a failure"; positive and negative items are alternated throughout the instrument. Response options are 'strongly agree', 'agree', 'disagree', and 'strongly disagree' (Ciarrochi & Bilich, 2006).

The instrument was originally designed as a Guttman scale, in which responses to the original ten statements are evaluated against six 'Scale Items', from which a seven-point refinement is produced to gain a uni-dimensional continuum. A higher score indicates lower self-esteem (Rosenberg, 1989).

For positively worded items, low self-esteem responses are Disagree and Strongly Disagree. For negatively worded items, low self-esteem responses are Agree and Strongly Agree (Rosenberg, 1989). The scale items are scored 'positively' if a minimum number of low self-esteem responses is met or exceeded.

Other design features were ease of administration, economy of time and face validity (Blascovich & Tomaka, 1991; Rosenberg, 1989). The RSES does not require special apparatus or

individual administration, it can be completed in just a few minutes, and the items openly and directly ask about self-esteem (Rosenberg, 1989).

Despite its design, common usage of the instrument typically involves awarding points to each response option (the number and scoring of which are often varied) and deriving a total Self-esteem score by summing the scores for each item (negative items are scored in reverse). Using this approach, a higher score is indicative of higher self-esteem (Blascovich & Tomaka, 1991; Gray-Little, Williams, & Hancock, 1997; Rosenberg, 1965). Use of the RSES in this way has occurred since at least 1972 (Yancey, Rigsby, & McCarthy, 1972) with little to no discussion or analysis of this change in methodology.

Both methods of scoring the Rosenberg Self-esteem Scale imply that the construct (self-esteem) is assumed to be quantitative, and the common approach implies that item responses represent interval-level data.

What the RSES measures

Rosenberg identifies self-esteem as an evaluation of the self-concept, according to what is most important to the individual (Rosenberg, 1986), and as a dimension of self-concept (Rosenberg, 1989).

In Black and White Self-Esteem: The Urban School Child, Rosenberg acknowledges the inconsistency with which self-terms are used, and reiterates his view of the construct of self-esteem (the object of measurement in the RSES), as detailed in Society and the Adolescent Self-Image, the account of the New York State study (Rosenberg, 1971, 1989).

In both texts, Rosenberg's definition is founded upon the perspective that global self-esteem is an attitude toward the self as a whole.

Rosenberg, Schooler, Schoenbach and Rosenberg (1995) clarify the conceptual difference between global and specific self-esteem as that of an attitude towards oneself as a whole versus an attitude towards a facet of oneself. They also ascribe the two elements of attitudes – affect and cognition - to global and specific self-esteem correspondingly, on the basis of evidence for global self-esteem being more related to wellbeing, and evidence for specific self-esteem being more related to behaviour.

According to Rosenberg's definition, high self-esteem refers to finding oneself 'good enough' rather than 'very good', or good in comparison to others; therefore the standard the self is being compared to is an internal one. High self-esteem implies feelings of self-worth and self-

respect, but not superiority over others. It implies self-acceptance, but with an additional desire to grow and improve.

Low self-esteem includes self-rejection, self-dissatisfaction and self-contempt, along with a lack of self-respect and dissatisfaction with the self.

Test development

The RSES was first used in 1965 in Rosenberg's New York State study, with a sample of 5024 high school students (Gray-Little et al., 1997; Rosenberg, 1989).

Rosenberg conducted pretesting prior to the New York State study, validating scores on the instrument against measures of depression via external observer ratings of 'often gloomy' and 'frequently disappointed', and another Guttman scale measuring "depressive affect" (Rosenberg, 1989). He conducted cross-validation against a measure of "neuroticism" that assessed psychosomatic symptoms, as well as self-reports referring to a list of 'psychogenic ailments'. Rosenberg also validated the RSES against peer-group reputation by correlating results with data on peer-reported classroom leadership, and active participation in class discussions.

In Rosenberg's original publication regarding the RSES, based on data from the New York State Study, he reported 'reproduceability' of 92%, and 'scalability' of 72%.

Since the New York State study, the instrument has been widely used for both adolescent and adult populations and has the distinction of being used as a comparison against which to evaluate new instruments (Blascovich & Tomaka, 1991). In 2003, Whiteside-Mansell and Corwyn found evidence for the comparable use of the RSES with adolescents and adults.

Psychometric properties

The PsycTESTS database record calls the RSES "the standard measure of self-esteem in psychological research" (Rosenberg, 1965).

There are a multitude of studies that have examined the psychometric properties of the RSES; what follows is a selection only.

Reliability

In their review of the RSES, Blascovich and Tomaka (1991) summarise the results of several studies – these demonstrate high internal consistency and test-retest reliability.

Gray-Little, Williams and Hancock (1997) conducted analysis of the RSES with 1,234 students. Their results included internal reliability of .88 and standard error of measurement of .23 on a (average) summed-score scale from 0 to 5.0. They concluded that the RSES is highly reliable and internally consistent, and deserving of its widespread use.

Bagley, Bolitho and Bertrand (2007) found the RSES to be internally reliable (α = .85-.90) in their study of 2,108 Canadian students.

Schmitt and Allik (2005) analysed data from a large scale study that included translations of the RSES into 28 languages with 16,998 participants across 53 countries. They concluded that the RSES is psychometrically sound with good internal reliability – the mean alpha across the 53 countries was .81, and the overall Guttman split-half reliability was .73.

Shapurian, Hojat and Nayerahmadi (1987) studied the psychometric properties of a Persian translation of the RSES with two samples of (537 total) Iranian students and found the measure to be highly reliable and stable over time. The alpha reliabilities for the scale were .82 and .83 for the two samples. Test-retest reliability was measured for 29 participants after a three-week interval, with an alpha of .74.

Vallieres and Vallerand (1990) studied the psychometric properties of a French translation of the RSES, and found internal reliability coefficients of .70, .83 and .88 across two samples (with one sample measured twice) and test-retest reliability after three weeks of .84 ($p < .001$).

Validity

In their study of 2,108 Canadian students, referenced above, Bagley, Bolitho and Bertrand (2007) found the RSES to be a valid instrument, showing construct validity with significant correlations ($p < .01$) in all age groups when correlated with the somatic problems and emotional disorder Ontario CHS scales, and the McMaster family relationships scale.

Griffiths, et al. (1999) found that the RSES was a significant predictor for each of their criterion variables, and that it had greater construct validity with a sample of 117 dieting disordered patients than another self-esteem scale.

Shapurian, Hojat and Nayerahmadi (1987), referenced above, reported a range of findings in support of the concurrent validity of the Persian translation of the RSES.

Hagborg (1993) reported strong concurrent validity via correlation between self-esteem as measured by the RSES and Global Self-Worth (RSE *r*(73) = .75, 56 of the variance).

As noted previously, the RSES was purposely designed as a Guttman scale to ensure a uni-dimensional scale. However, many studies have examined the items within the Rosenberg Self-Esteem Scale, in order to determine whether they truly represent a uni-dimensional construct. A variety of researchers have found evidence for either one or two factors within the items, resulting in a minor controversy within the literature (Hagborg, 1993).

Kohn and Schooler (1969) were perhaps the first to propose two factors, arguing via factor analysis that the positively and negatively worded items within the RSES represent unique factors. They labeled these self-confidence and self-deprecation respectively.

Goldsmith (1986) proposed that not only was self-esteem as measured by the RSES multidimensional, but that the constructs measured were different depending on the population being measured.

Conversely, Gray-Little et al., (1997; using the RSES as a Likert scale with five response options per item) found that their data supported Rosenberg's uni-dimensional construct, with a slight increase in reliability using a weighted scoring approach.

Greenberger, Chen, Dmitrieva & Farruggia (2003) found support for the uni-dimensionality of the RSES and proposed that the two-factor structure of the RSES is an artifact of positive and negative item-wording.

Several other authors found support for the RSES measuring a single construct, using translations of the RSES into Estonian (Pullmann & Allik, 2000), Persian (Shapurian et al., 1987) and French (Vallieres & Vallerand, 1990).

The full scale of the RSES can be retrieved from PsycTESTS: doi: 10.1037/t01038-000.

Occupational Personality Questionnaire (OPQ)

The personality assessment that was completed by participants in both Sub-projects A and B was the OPQ-32R.

Description

The Occupational Personality Questionnaire is a widely used and well-regarded measure of work-related personality. The original version of the OPQ, published in 1984 (Saville, Holdsworth, Nyfield, Cramp, & Mabey, 1984) provided a groundbreaking move away from existing clinical tests via an assessment designed for organisational/occupational use (British Psychological Society, 2009).

There are now a number of versions of the OPQ, which have varying formats and scoring methodologies.

The OPQ-32 version is available in three formats: OPQ-32N (normative version), OPQ-32I (ipsative version) and most recently, OPQ-32R (forced-choice version; A. Brown & Bartram, 2009b)[14].

The OPQ-32R is comprised of 104 blocks of three statements; for each block the test-taker must assign the descriptors 'Most like me' and 'Least like me' to one of the statements. The assessment is not timed, but typically takes 30 minutes to complete (Leung & Porchea, 2014). Comparatively, the OPQ-32N is comprised of 230 items with a 5-point Likert scale for each and takes approximately 35 minutes to complete, and the OPQ-32I is comprised of 104 blocks of four items and takes approximately 45 minutes to complete (Leung & Porchea, 2014).

While there are pen and paper options for completing and scoring other versions, the OPQ32r must be completed and scored on a computer, via web-based application (British Psychological Society, 2009).

What the OPQ-32R measures

The OPQ-32R is a mainstream, trait-based measure of work-related personality.

The model of personality it is built upon was created via its own development, such that the scales represent the operational definition of the model (Barrett, Kline, Paltiel, & Eysenck, 1996; Saville, Sik, Nyfield, Hackston, & MacIver, 1996). Burke describes the OPQ as a transparent instrument, that "does not propose a rigid or deterministic view of personality,

[14] Note that ipsative and normative in this instance refer to questionnaires in which the sum of scales measured for each respondent is equal, e.g. via selecting most and least applicable from four items (ipsative) or unequal, e.g. selecting from 1-7 I relation to each item's applicability (normative). See Saville and Willson (1991).

but the model underlying it does propose that current and future behaviour is influenced by the personality of individuals and work groups" (2008, p. 88).

The OPQ-32R (in contrast to earlier versions and formats) uses Item Response Theory (IRT) to provide scores along 32 scales, or "dimensions of people's preferred style of behaviour at work" (A. Brown & Bartram, 2009b, p. 2). The scales are straightforward, with scale items reasonably and transparently linked to each scale (Psychological Testing Centre, 2016).

The 32 scales are organised into three primary domains as follows (Leung & Porchea, 2014):

- Relationships with People: Persuasive, Controlling, Outspoken, Independent-Minded, Outgoing, Affiliative, Socially Confident, Modest, Democratic, and Caring
- Thinking Style: Data Rational, Evaluative, Behavioral, Conventional, Conceptual, Innovative, Variety Seeking, Adaptable, Forward Thinking, Detail Conscious, Conscientious, and Rule Following
- Feelings and Emotions: Relaxed, Worrying, Tough Minded, Optimistic, Trusting, Emotionally Controlled, Vigorous, Competitive, Achieving and Decisive.

The OPQ32 also measures a fourth domain, Dynamism (Vigorous, Achieving and Competitive), which relates to sources of energy (A. Brown & Bartram, 2009b). All versions of the OPQ additionally include a Social Desirability or Consistency scale to detect 'faking good' (Leung & Porchea, 2014).

The 32 scales can also be categorised into eight categories: Influence (Persuasive, Controlling, Outspoken and Independent Minded), Sociability (Outgoing, Affiliative and Socially Confident), Empathy (Modest, Democratic and Caring), Analysis (Data Rational, Evaluative and Behavioral), Creativity and Change (Conventional, Conceptual, Innovative, Variety Seeking and Adaptable), Structure (Forward Thinking, Detail Conscious, Conscientious and Rule Following), Emotion (Relaxed, Worrying, Tough Minded, Optimistic, Trusting and Emotionally Controlled) and Dynamism (Vigorous, Competitive, Achieving and Decisive; SHL Limited, 2013).

The OPQ Profile report (one of many report outputs from the OPQ32) presents the scores for each scale alongside each of the three primary domains as well as the eight categories (SHL Limited, 2013).

The scales of the OPQ32 can also be mapped onto the Five-factor Model, though the OPQ32 claims to measure a wider domain of personality than just the FFM (Bartram, 2005; A. Brown & Bartram, 2009a). In their discussion of how the OPQ32 relates to the Five-factor Model,

Bartram and Brown (2005) refer to the 32 scales as first-order personality traits that underlie the second-order factors of the FFM.

Test development

The framework of personality underlying the OPQ32 was developed from existing personality theories and assessment tools, work-related information and feedback from organisations, as well as employee-generated repertory grid data (Barrett et al., 1996; Leung & Porchea, 2014; Stanton, 1994). Using this data, hundreds of test items were created and tested across multiple organisations, resulting in an initial model with 40 scales (Barrett et al., 1996; Leung & Porchea, 2014).

Test development began in 1981, with the first phase concluding in 1984 (Saville et al., 1996). By the end of phase one, the original 40 scales had been reduced to 32 and then to 30, as items relating to the scales were tested on hundreds of participants (Leung & Porchea, 2014). During the second phase of development, from 1984 to 1994, items were refined, scales were mapped to the FFM, and cross-cultural differences were investigated, among other analyses. The final phase of development took place from 1994 to 1999, during which the OPQ32 was launched and the OPQ-32N and OPQ-32I were published (Leung & Porchea, 2014). The OPQ-32R was released in 2009; the reasons behind its development and the improvements it provides are clearly described in a dedicated supplement to the OPQ32 Technical Manual (A. Brown & Bartram, 2009a).

There are 86 norm groups for the OPQ-32N and OPQ-32I versions. The OPQ-32R norms were developed via application of Item Response Theory (IRT) scoring to OPQ-32I responses, and the technical manual states that all 32I population, user and local norms can be used with the 32R (Leung & Porchea, 2014).

Psychometric properties

In addition to the SHL OPQ representing a globally pioneering effort via the creation of the first occupational personality inventory, the majority of critical reviews have been favourable. Specifically, the OPQ32 is acknowledged as a sound test by reviewers in the Mental Measurements Yearbook series and the British Psychological Society (BPS) Psychological Test Centre (Leung & Porchea, 2014; Psychological Testing Centre, 2016). The BPS praises the

OPQ32 stating it is “at the top of the first rank of personality tests" (Psychological Testing Centre, 2016, p. 26).

While numerous validation studies of the OPQ32 are known to SHL consultants, the outcomes mostly remain the property of SHL. The published studies citable in this thesis most often relate to other versions of the OPQ.

Prewett, Tett and Christiansen (2013) chose to review the OPQ-32N and OPQ-32I among 12 personality inventories, noting that OPQ-32R draws from the same item pool. Joubert, Inceoglu, Bartram, Dowdeswell and Lin (2015) conducted a study comparing the OPQ-32N and OPQ-32R. Their results showed support for the equivalence of the data structures, measurement precision, and scaling properties of these versions.

Reliability

The OPQ-32R is one of a handful of personality inventories currently benefiting from a development process utilising Item Response Theory (Embretson & Reise, 2013; Hutchinson, 1991). One of the specific advantages is the improved reliability of OPQ trait scales.

Brown and Bartram (2009a) report a composite reliability score for OPQ-32R scales of .84, compared to a median Cronbach's alpha of .81 for the OPQ-32I.

Hendy (2017) examined the reliability of Big Five personality scales derived from the OPQ-32R. She found internal consistency coefficients ranging from .52 to .89.

Validity

Saville, Sik, Nyfield, Hackston and MacIver conducted two validation studies encompassing 710 managers in the UK. Results from the OPQ CM3+5 and OPQCM4.2 showed, “predictable, significant and substantial correlations with criteria of management job success” (1996, p. 260).

Robertson and Kinder (1993) conducted a meta-analysis of 20 validation studies involving the OPQ, with regard to criterion-related validity of personality variables. They found that personality scales demonstrated criterion-related validity, and that this validity was incremental over and above cognitive ability test results. (See also, Salgado, 1996).

Matthews, Stanton, Graham and Brimelow (1990) analysed the factor structure of the OPQ Concept 5 version and concluded that the OPQ corresponded to the five factors of the FFM, as

well as additional reliable variance, supporting the idea that the OPQ measures a wider scope of personality than the FFM. They found support for a 21-factor model, however, rather than the 31 scales of the Concept 5. Matthews and Stanton (1994), and Barrett, Kline, Paltiel and Eysenck (1996) also analysed the factor structure of the OPQ Concept 5.2 Model and found that the 31 scale structure was not supported by their data.

Ferguson, Payne and Anderson (1994) conducted a factor analysis of the FMX5-Student version of the OPQ. They concluded that the OPQ scales represented five higher order factors that resembled the Big Five.

4.1.3 Strategy for quantitative analysis

As stated in the Hypotheses chapter, Hypothesis 1 predicts that completion of a personality assessment, as part of a selection process, will be associated with a change in self-concept (including components thereof), and/or self-esteem.

Hypothesis 1a predicts that any/all of the following variables may influence this change: “No. of roles applied for in last month”, “expectation of personality assessment as positive/negative experience”, “confidence in correct interpretation of personality assessment results”, “confidence in correct interpretation of ability assessment results”, sex, ethnicity, age, “current employment status”, “reason for applying for role”, “expected salary”, “previous experience of personality assessment(s)”.

The strategy for testing these hypotheses is twofold, with checks on the psychometric properties of the data collection instruments preceding hypothesis testing proper.

To ascertain whether there has been any change to self-perception, Hypothesis 1 will be tested via t-tests for dependent means.

To discover whether any change has been influenced by other variables (Hypothesis 1a), bivariate correlations will be generated and MANOVAs conducted.

4.1.4 Checking psychometric properties

While complete re-validation was beyond the logistic possibilities of this thesis, one requirement is to re-examine the psychometric properties of all imported measuring instruments for local use. The following sub-sections report the outcomes of such checks. All three elements of the ‘questionnaire’ (i.e. the merged data collection tool) were examined.

Researcher-created questionnaire items

As described above, the two merged data collection instruments ('questionnaires') completed by participants in Sub-project A contained a number of (researcher-created) items in addition to the SFSCS and RSES scales[15]. The intent of these items was to cast a net of additional information gathering in line with the exploratory nature of this research.

Researcher-created items within the first questionnaire were categorised as follows:

- Context of job application
- Expectations regarding psychometric assessments within the selection process
- Previous experience of personality assessments

Researcher-created items within the second questionnaire were categorised as follows:

- Experience of completing personality assessment
- Experience of receiving results/feedback (not completed by any participants)
- Impact of completing personality assessment
- Outcome of job application

Frequency tables for these items, within their categories, are included in Appendix 3.

<u>Descriptive statistics</u>

Descriptive statistics for interval-level researcher-created items are included in Appendix 4.

Correlations between interval-level items (within item categories) were also calculated (Appendix 5).

- The two interval-level items within the category Context of job application ("no. of roles applied for in last month" and "expected salary"), showed a statistically significant correlation, $r(25) = .518$, $p < .01$.
- Two of the items within the category Expectations regarding psychometric assessment within the selection process ("confidence in correct interpretation of personality assessment results", and "confidence in correct interpretation of ability assessment results"), also showed a statistically significant correlation, $r(25) = .599$, $p < .01$.

[15] There were additional items in both questionnaires that served only to validate participant responses. E.g. "Please enter the name of the organisation from which you received the link to this questionnaire." These items were excluded from all analyses.

These results could signify that for this participant group, psychometric ability assessments and personality assessments were viewed as having similar validity.

Two pairs of items within the category Experience of completing personality assessment also showed a statistically significant correlation:

- "Experience of personality assessment as positive/negative" showed a statistically significant correlation with "confidence that personality assessment results were interpreted correctly", $r(25) = .596$, $p < .01$.
- "Confidence that personality assessment results were interpreted correctly" showed a statistically significant correlation with "personality assessment allowed true and complete picture", $r(25) = .616$, $p < .01$.

Reliability

Cronbach's alpha was calculated for the interval-level questions within the categories Expectations regarding psychometrics within selection process and Experience of completing personality assessment (Appendix 6).

Two items were reverse scored for the purpose of calculating reliability ("Expectation of personality assessment as positive/negative experience" and "experience of personality assessment as positive/negative"). Both sets of items showed good reliability (.669 and .766 respectively).

4.1.5 Partial re-validation of SFSCS and RSES

Because the SFSCS and RSES were not developed using New Zealand participants, their appropriateness cannot be assumed for use with a New Zealand participant group.

Whenever a psychometric tool is imported into a new context, its properties in that context should be re-examined. This includes checks of administrative viability, cultural appropriateness, and participant reactions, as well as revalidation in terms of reliability, and content-, criterion- and construct-related validity (Arnold & Smith, 2013; Cronbach & Drenth, 1973; Dana, 1993).

The following attempts were made to validate the use of these instruments in the New Zealand context, however time constraints made a comprehensive analysis impossible.

Where possible, factor structure has been compared to original data, and descriptive statistics, reliability, and convergent validity (correlating the two measures) have been calculated.

Both instruments were developed in North America where there are cultural overlaps with New Zealand culture; both instruments were developed in English.

For the sake of completion of the project, and based on the following analyses, it was assumed that both instruments were suitable for use with New Zealand participants.

SFSCS

Descriptive statistics

Descriptive statistics for each subscale measured by the SFSCS were calculated for the participants in Sub-project A, before and after completing the personality assessment (Appendix 7).

Subscales were scored by summing response scores for items within each subscale. Power and Gifted subscales were weighted by .857 and 1.2 respectively, for the purpose of subscale comparison. A Total Self-concept score was derived by adding the subscales Likeability, Task Accomplishment, Power (weighted), Gifted (weighted) and Moral, and subtracting the subscale Vulnerability, as per Stake (1994).

Means for the SFSCS subscales varied before and after completion of the personality assessment, but retained the same ranking (in decreasing order): Moral, Task Accomplishment, Likeability, Gifted, Power and Vulnerability.

The Gifted subscale had the greatest range, largest standard deviation and largest variance, both before and after completion of the personality assessment.

The Moral subscale had the greatest absolute skew before completion of the personality assessment, and Task Accomplishment had the greatest absolute skew after completion of the personality assessment.

Power had the greatest absolute kurtosis before completion of the personality assessment, and Task Accomplishment had the greatest absolute kurtosis after completion of the personality assessment.

Reliability

Cronbach's alpha was calculated for all subscales of the SFSCS as well as the total scale, for the participants in Sub-project A before completing the personality assessment (Appendix 8). Vulnerability subscale item scores were reversed for the purpose of calculating reliability for the total scale.

The total scale demonstrated high reliability ($\alpha = .878$), with six items ("Hard worker", "Friendly", "Forceful", "Easily hurt", "Self-conscious", "Pleasant") that would have improved the reliability of the scale if they had been removed.

All subscales (apart from Vulnerability ($\alpha = .683$)) demonstrated high reliability, with alpha coefficients ranging from .802 (Power) to .893 (Gifted). For Likeability and Power, there were no items that would have improved the reliability if they had been removed.

For the Task Accomplishment subscale one item ("Can concentrate well on a task") would have improved the reliability of the subscale if they had been removed. For the Vulnerability subscale two items ("Lacks confidence", "Self-conscious") would have improved the reliability of the subscale if they had been removed. For the Gifted subscale one item ("Creative") would have improved the reliability of the subscale if they had been removed. For the Moral subscale one item ("Law-abiding") would have improved the reliability of the subscale if they had been removed.

Validity

As a limited attempt to locally re-validate for construct validity, checking the component/ factor structure of SFSCS over New Zealand data appeared to be a promising avenue. Principal Component Analysis (PCA) is a commonly used method to supply evidence for the construct validity of a test (Bryant & Yarnold, 1995). In this case, the evidence had to be limited to correlation matrices because the actual PCA was unviable in Sub-project A due to small N relative to the number of items in the test.

In validating the SFSCS, Stake (1994) theorised that the subscales of the SFSCS should be correlated more highly with a measure of self-esteem than with social desirability. Her results supported this theory, with a high correlation between the SFSCS and RSES (for Total Self-concept, $r = .62$, $p < .0001$).

Using Sub-project A data, a correlation matrix including the subscales of the SFSCS and the RSES was produced. Separate matrices were generated from data collected before and after completion of the personality assessment (Appendix 9).

Comparable to Stake's results, the correlation between Total Self-concept and Self-esteem Sub-project A (before completing the personality assessment) was (r25) = .755 ($p < .01$) and the median subscale correlation was .39. After completing the personality assessment, the correlation between Total Self-concept and self-esteem was r(25) = .808 ($p < .01$) with a median subscale correlation of .47.

These correlations provide indirect support for the validity of the SFSCS with the participant group for Sub-project A, via a similar relationship between self-concept and self-esteem.

Face validity was also checked and 20 of 27 participants affirmed that they thought the personality assessment was relevant to their job application.

RSES

Descriptive statistics

Descriptive statistics for the RSES were calculated for the participants in Sub-project A, before and after completing the personality assessment (Appendix 10).

The RSES was scored by summing response scores for all items (with negative items scored negatively) to achieve a total Self-esteem score.

The mean showed a small increase after completion of the personality assessment, however the standard error also increased. The range, standard deviation, variance, skew and kurtosis were all greater after completion of the personality assessment.

Reliability

Cronbach's alpha was calculated for the RSES (before completing the personality assessment; Appendix 11.

The total scale demonstrated high reliability ($\alpha = .877$), with two items ("I am able to do things are well as most other people", "I feel I do not have much to be proud of") that would have improved the reliability of the scale if they had been removed.

Validity

While a complete validation study for construct validity was beyond the scope of this thesis, an attempt was made to clarify whether New Zealand data would "reproduce" the component structure. Instead of exploratory factor analysis, Principal Component Analysis (PCA) was conducted on the 10 items of the RSES (for the participants in Sub-project A, before completing the personality assessment; Appendix 12). Orthogonal rotation (VARIMAX) was chosen (Bryant & Yarnold, 1995; Field, 2009).

While N is low, the Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .702. All KMO values for items were > .55. Bartlett's test of sphericity $\chi 2(45) = 150.864$, $p < .000$, indicated that correlations between item variables were sufficiently large.

The nature of components as well as eigenvalues were reviewed. Three components had eigenvalues over Kaiser's criterion of 1.0. This combination explained 74.35% of the variance.

4.2 Sub-project A: Results

4.2.1 Results of testing

Following on from analyses of the psychometric properties of the data collection instruments, formal testing of the below hypotheses (including pre-test checks) was conducted.

Hypothesis 1 states that completion of a personality assessment as part of a selection process will be associated with a change in:

- Any or all of the following components of self-concept as measured by the Six-Factor Self-Concept Scale: Likeability, Task Accomplishment, Power, Vulnerability, Gifted, Moral; and/or,
- Total Self-concept as measured by the Six-Factor Self-Concept Scale; and/or,
- Self-esteem as measured by the Rosenberg Self-Esteem Scale

Hypothesis 1a states that positive or negative change to self-concept and/or components of self-concept and/or self-esteem will be influenced by any of the following variables:

- No. of roles applied for in last month
- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results
- Confidence in correct interpretation of ability assessment results
- Sex

- Ethnicity
- Age
- Current employment status
- Reason for applying for role
- Expected salary
- Previous experience of personality assessment(s)

Calculation of change scores

A change score (shift score, gain score) for each of the self-perception variables was calculated by subtracting the self-concept/self-esteem scores yielded by Questionnaire 1 from those yielded by Questionnaire 2. Such a change score is intended to quantify shift over time (Bereiter, 1963; Cronbach & Furby, 1970).

The use of simple change scores, especially in terms of measurement error and reliability, has been the target of critical scrutiny, with some authors less accepting than others (Burr & Nesselroade, 1990). Detail regarding the theoretical basis, advantages, and limitations of change scores will be revisited in the Discussion chapter.

Analysis of distributions

As discussed above, variables were checked for the properties of their distributions (see Appendices 7 and 10). For each variable, indices of central tendency (mean, median, mode), spread (variance and SD), as well as skewness and kurtosis statistics were generated.

The dependent variables in relation to self-perception had distributions with skewness ranging from -1.064 to .563 before completion of the personality assessment. The same indicators ranged from -2.471 to .206 after completion. Kurtosis ranged from -1.120 to .695 before completion of the personality assessment, and from -1.293 to 9.803 after completion. Such degrees of skewness and kurtosis are commonly estimated for psychometric score variables with a small N and non-representative samples.

Distributions of the change scores were also examined. Likeability and Moral subscales as well as Total Self-concept were more skewed than the other change score variables (Appendix 13). Moral and Total Self-concept also showed significant kurtosis.

Kolmogorov-Smirnov tests were then performed to see the extent to which these variables deviate from symmetric distributions (Appendix 14). Among the change score variables, Task Accomplishment, Power, Vulnerability, Giftedness and Self-esteem had distributions closest to symmetric.

Change in self-perception over time

To test Hypothesis 1, a paired-samples t-test was conducted to compare self-concept subscales, Total Self-concept and Self-esteem scores before and after completion of the personality assessment (Appendix 15).

Only the Power (weighted) component of self-concept showed a significant effect.

On average, participants' weighted Power scores after completing the personality assessment (M = 23.02, SE = .97) were lower than before completing the personality assessment (M = 24.54, SE = .90). This difference, 1.524, BCa 95% CI [.59, 2.46], was significant $t(26) = 3.34$, $p = .003$, and represented a small-sized effect, $r = .30$.

Using a Bonferroni correction (Bonferroni, 1935, 1936) for repeated t-tests, $p < .05$ was replaced by the more conservative requirement $p < .006$. At this level, the effect for Power remains significant.

The direction of potential shift, i.e. increase versus decrease of score after completion of the personality assessment, was also of interest. While not all t values produced were significant, demonstrated tendencies appear to be parallel. All subscales of self-concept, as well as Total self-concept showed a decrease in means following completion of the personality assessment.

Variables influencing change in self-perception

To partially test Hypothesis 1a, correlations were calculated between the variables below and each of the change score variables in order to assess whether these interval-level variables may have influenced any change to self-perception. None of these correlations were significant.

- No. of roles applied for in last month
- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results
- Confidence in correct interpretation of ability assessment results

For a multivariate approach to testing Hypothesis 1a, several analyses of variance were conducted via MANOVA. In the output, Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root were evaluated (Field, 2009).

Change score variables for self-perception served as dependent variables. Specifically, the self-concept subscales of the SFSCS (Likeability, Task Accomplishment, Power, Vulnerability, Moral, Gifted), Total Self-concept (measured by the SFSCS), and Self-esteem (measured by the RSES) were the dependent variables.

Independent variables were:

- MANOVA 1 independent variables: sex, age group, ethnicity
- MANOVA 2 independent variables: "Current employment status", "reason for applying for role", "expected salary"
- MANOVA 3 independent variables: "Previous experience of personality assessment(s)", "reason(s) for prior completion of personality assessment(s)"

MANOVA 1 did not demonstrate any significant effect of sex and/or ethnicity and/or age on the dependent variables. Separate ANOVAs on the independent variables revealed no significant F value representing any effects on the dependent variables.

Using Roy's largest root, MANOVA 2 demonstrated a significant effect of "reason for applying for role" on the dependent variables, $\Lambda = 15.49$, $F(4,3) = 11.62$, $p < .05$. There was also a significant effect of "reason for applying for role" + "expected salary" on the dependent variables, $\Lambda = 17.65$, $F(3,3) = 17.65$, $p < .05$; Appendix 16).

However, separate ANOVAs on the independent variables revealed no significant F value representing any effects on the dependent variables.

MANOVA 3 did not demonstrate any significant effect of "previous experience of personality assessment(s)" and/or "Have you previously completed a personality assessment as part of a job application?" and/or "Have you ever completed the OPQ?" on the dependent variables. Separate ANOVAs on the independent variables revealed no significant F value representing any effects on the dependent variables.

4.2.2 Intercorrelation of self-perception variables

The variables self-concept and self-esteem were expected to be associated with each other. In fact, as discussed above, correlation between the SFSCS and the RSES was counted as evidence for the convergent validity of the SFSCS (Stake, 1994; Appendix 9).

There were significant correlations between Self-esteem and the following subscales of self-concept within the data collected before completion of the personality assessment: Task Accomplishment (.456), Power (.520), Vulnerability (-.629), and Gifted (.551).

In the data collected after completion of the personality assessment, Self-esteem significantly correlated with the following subscales of self-concept: Likeability (.547), Task Accomplishment (.740), Vulnerability (-.719), Gifted (.543) and Moral (.394).

Change in Self-esteem correlated significantly with Change in Vulnerability (-.564).

The only self-concept subscale change scores to correlate significantly with each other were Change in Likeability and Change in Vulnerability (-.453).

See Appendix 17 for correlations between change scores for the SFSCS and RSES.

5 Sub-project B

5.1 Sub-project B: Method

5.1.1 Participants

Participant source

Participants were Massey University students enrolled in papers administrated by participating lecturers during the period July – September 2015, who received and responded to a research invitation posted on the webpage for their paper(s), or received via email from a participating lecturer.

Method of recruitment

Lecturers from the Massey University School of Psychology, School of Engineering and Advanced Technology, and School of Management were approached by the researcher via email and asked to post an invitation to participate on the website for their paper(s) so that students in their papers would see it.

Some lecturers also advertised the study to their students via email and word of mouth.

Participant characteristics

See Tables 6-10 for distributions of participants in relation to sex, age group, ethnicity, and previous experience (including reasons for completion) of personality assessments.

Table 6: Sex – Sub-project B

Sex	Count	Percentage
Female	82	89.13
Male	10	10.87

Table 7: Age group – Sub-project B

Age group	Count	Percentage
Under 16	0	0.00

16 to 24 years	30	32.61
25 to 34 years	29	31.52
35 to 44 years	18	19.57
45 to 54 years	12	13.04
55 to 64 years	2	2.18
65 years and over	1	1.08

Table 8: Ethnicity – Sub-project B

Statistics New Zealand Level 1 classification[16]	**Count**	**Percentage**
European	80	86.96
Maori	7	7.61
Pacific Peoples	0	0.00
Asian	5	5.43
Middle Eastern	0	0.00
Latin American	0	0.00
African	0	0.00

Table 9: Previous experience of personality assessment(s) –Sub-project B

Previous experience of personality assessment(s)	**Count**	**Percentage**
Yes	44	47.83
No	48	52.17

Table 10: Reason(s) for prior completion of personality assessment(s) – Sub-project B

Reason	**Count**	**Percentage**
As part of a job application	17	38.64
Personal development (at work)	15	34.10
Personal development (outside work)	17	38.64
In school/university setting	19	43.18

[16] Ministry of Health (2004)

The most commonly completed personality assessment (of the 14 specifically queried) was the Myers-Briggs Type Indicator (MBTI; version not specified) with a frequency of 80 completions across 33 participants. The second most frequently completed assessment was the 16 Personality Factors Questionnaire (16PF) with 13 completions across 11 participants, and the third most frequently completed assessment was the Occupational Personality Questionnaire (OPQ; version not specified) with 10 completions across 10 participants.

Exclusionary criteria and number of participants excluded

Only current Massey University students were invited to participate, and any participant who did not complete both questionnaires as well as the personality assessment was excluded.

Of the 144 students who accepted the research invitation, 14 were excluded because they did not complete Questionnaire 1. Of the 130 students who were sent the personality assessment, 31 were excluded because they did not complete the personality assessment. Of the 99 students who were sent the link to Questionnaire 2, seven were excluded for non-completion.

Whether participant group was representative

Sub-project B did not include a process of sampling from a population, let alone representative sampling (Field, 2009; Frankfort-Nachmias, 1996; Hansen et al., 1993).

Representativeness is examined in detail in the Discussion chapter.

Acceptance rates for participants who were solicited

Because the invitation to individual participants was an online advertisement on the webpage for specific papers, it was not possible to determine the exact proportion of students who accepted the invitation (to participate) out of the total number of those who were privy to the invitation.

How participants were assigned to conditions

There was only one condition in this study, with all participants going through the same process. The study was an observational study; the researcher did not assign participants to the condition, rather participants self-selected to participate or not participate.

Participants were offered the chance to complete the OPQ-32R, a well-known and widely used personality assessment, and to receive a written feedback report (representing a rare opportunity). Participants who completed both questionnaires and the personality assessment were also entered into a draw with five chances to win a $100 Prezzy Card.

5.1.2 Data collection

Contacting university lecturers

After gaining organisational sponsorship for a study using Massey University students, contact was made with several Massey University lecturers, to request that they advertise Sub-project B to students in their classes. Several lecturers agreed to post a supplied advertisement on the webpage for each of their classes; a few lecturers also advertised the study to their students via email and word of mouth.

Design

As in Sub-project A, this study employed a repeated-measures design, comparing measures of individuals' self-perception before and after completion of a personality assessment and receipt of a written feedback report. As detailed in the Hypotheses chapter, the original design of the study included additional waves of measurement, but had to be curtailed due to low participant commitment.

Repeated measures refers to the use of one participant group, all of whom experience both independent variable conditions (Alexander & McGarty, 2014).

The within-subjects factor had two levels:

- Before completing a personality assessment for selection
- After (i.e. within a few days of) completing a personality assessment (during which time participants also received and read a written feedback report regarding their personality assessment results).

This research design can be susceptible to order effects (also called carry-over effects), including practice effects, fatigue effects, history effects and maturation (Nestor & Schutt, 2015; Wilson & Joye, 2017). The potential impacts of these are discussed in the Discussion chapter.

Sequence

The sequence for participants was as follows:

1. Each participant receives an online or email invitation to participate in the research project for five chances to win $100. The link to Questionnaire 1 is included in the invitation.

2. The participant completes Questionnaire 1.

3. Following completion of Questionnaire 1, each participant receives an email from SHL New Zealand with a link to an online personality assessment (OPQ-32R).

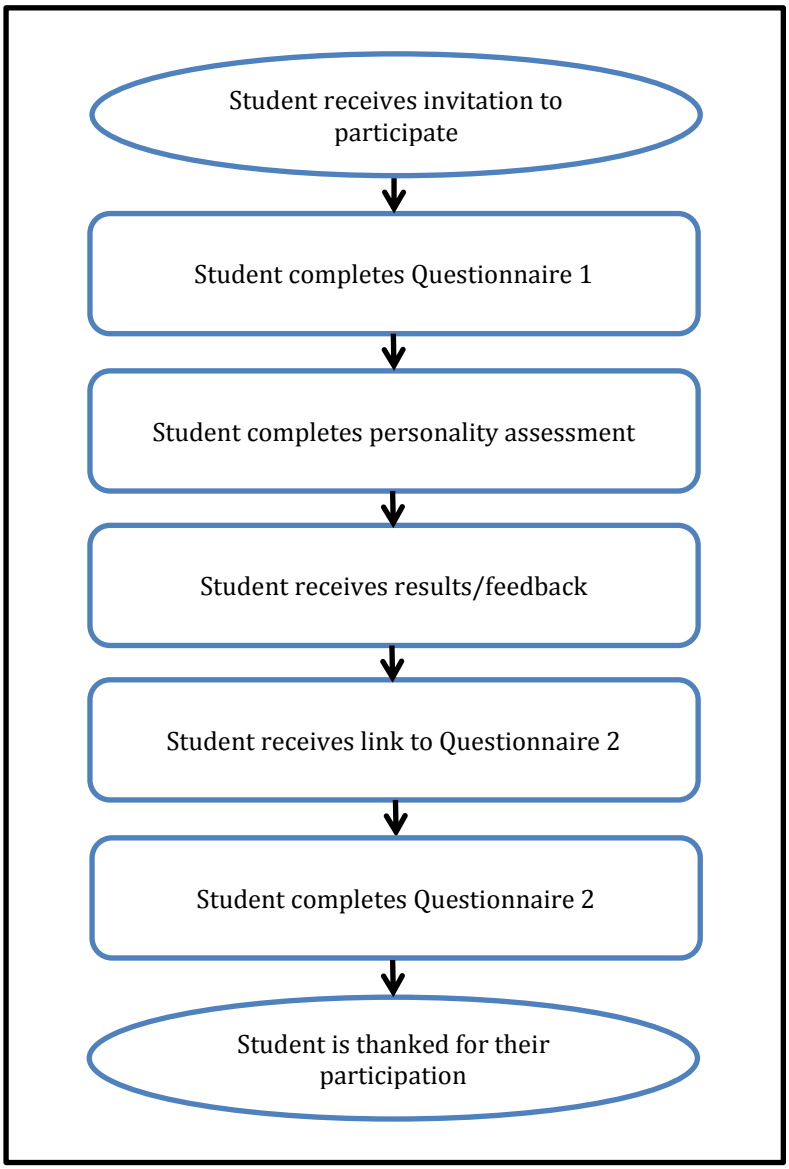
4. Following completion of the personality assessment, each participant receives an email from SHL New Zealand containing a descriptive summary of his or her personality assessment results.

5. Following receipt of the personality assessment results, each participant receives an email from the researcher to thank them for their participation and to request that they complete Questionnaire 2 as soon as possible. The link to Questionnaire 2 is included in this email.

6. The participant completes Questionnaire 2.

Instructions and order of events were the same for all participants. See Figure 3.

Figure 3: Sub-project B sequence



Instruments

Instruments for data collection used in this sub-project were two self-report 'questionnaires' administered via the online tool Qualtrics Surveys. A more accurate description is 'merged data collection instrument', however for ease of communications, these were labelled as questionnaires for university lecturers and individual participants.

Participant responses and associated metadata were downloaded from Qualtrics Surveys and imported into IBM SPSS.

Each ‘questionnaire’ comprised a range of forced-choice (categorical, nominal and ordinal level) items and free-form items used and/or created by the researcher, as well as full scales of the Six-Factor Self-Concept Scale and the Rosenberg Self-Esteem Scale.

Questionnaires and questionnaire items were administered in the same order for all participants. The variables measured within the questionnaires were as follows.

- Demographics – sex, age group, ethnicity
- Expectations regarding personality assessment – “expectation of personality assessment as positive/negative experience”, “confidence in correct interpretation of personality assessment results”
- Previous experience of personality assessment(s) – “previous experience of personality assessment(s) [Y/N]”, “reason(s) for prior completion of personality assessment(s)”, “specific tests completed”, ‘number of times completed”
- Self-concept (pre-assessment)
- Self-esteem (pre-assessment)
- Experience of completing personality assessment – “experience of personality assessment as positive/negative”, “difficulty completing personality assessment”, “personality assessment allowed true and complete picture”, “questions that did not apply”, “consistency of responses”
- Impact of completing personality assessment – “considered something new”, “thought differently about self”, “changed optimism”
- Experience of receiving feedback – “level of detail provided”, “easy to understand”, confidence that interpretation of results was correct”, “feedback as positive/negative experience”, feedback provided complete picture”, “feedback was surprising”
- Impact of receiving feedback – “feedback considered something new”, “feedback caused to think differently”, “would answer differently post-feedback”, “feedback of personal use”, “feedback changed optimism”
- Self-concept (post-assessment)
- Self-esteem (post-assessment)

Numeric values were assigned to response options for (non-freeform) researcher-created items within the questionnaires, according to the assumed scale of measurement for each item. Assignment of numeric values to response options for the Six-Factor Self-Concept Scale and Rosenberg Self-Esteem Scale is detailed in the sub-sections below.

Six-Factor Self-Concept Scale and Rosenberg Self-Esteem Scale

For a critical summary of the SFSCS and RSES, including validation and reliability research, see Chapter 4: Sub-Project A: Method and Results.

OPQ-32R

For a critical summary of the OPQ, including validation and reliability research, see Chapter 4: Sub-Project A: Method and Results.

5.1.3 Strategy for quantitative analysis

As stated in the Hypotheses chapter, Hypothesis 2 predicts that completion of a personality assessment and receipt of results/feedback will be associated with a change in self-concept (including components thereof), and/or self-esteem.

Hypothesis 2a predicts that any/all of the following variables may influence this change: “Expectation of assessment as positive/negative experience”, “confidence in correct interpretation of personality assessment results”, sex, age group, ethnicity,” previous experience of personality assessment(s)”, “reason(s) for prior completion of personality assessment(s)”

The strategy for testing these hypotheses is twofold, with checks on the psychometric properties of the data collection instruments preceding hypothesis testing proper.

To ascertain whether there has been any change to self-perception, Hypothesis 2 will be tested via t-tests for dependent means.

To discover whether any change has been influenced by other variables (Hypothesis 2a), bivariate correlations will be generated and MANOVAs conducted.

5.1.4 Checking psychometric properties

While complete re-validation was beyond the logistic possibilities of this thesis, one requirement is to re-examine the psychometric properties of all imported measuring instruments for local use. The following sub-sections report the outcomes of such checks. All three elements of the "questionnaire" (i.e. the merged data collection tool) were examined.

Researcher-created questionnaire items

As described above, the two merged data collection instruments ('questionnaires') completed by participants in Sub-project B contained a number of (researcher-created) items in addition to the SFSCS and RSES scales. As in Sub-project A, the intent of these items was to cast a net of additional information gathering in line with the exploratory nature of this research.

Researcher-created items within the first questionnaire were categorised as follows:

- Expectations regarding personality assessment
- Previous experience of personality assessments

Researcher-created items within the second questionnaire were categorised as follows:

- Experience of completing personality assessment
- Impact of completing personality assessment
- Experience of receiving feedback
- Impact of receiving feedback

Frequency tables for these items, within their categories, are included in Appendix 18.

<u>Descriptive statistics</u>

Descriptive statistics for interval-level researcher-created items are included in Appendix 19.

Correlations between interval-level items (within item categories) were also calculated (Appendix 20).

The two items within the category Experience of completing personality assessment ("experience of personality assessment as positive/negative" and "personality assessment allowed true and complete picture"), showed a statistically significant correlation, $r(90) = -.532$, $p < .01$.

Four pairs of items within the category Experience of receiving feedback showed a statistically significant correlation, as below.

- "Level of detail provided" showed a statistically significant correlation with "confidence that interpretation of results was correct", $r(90) = .350$, $p < .01$.
- "Level of detail provided" also showed a statistically significant correlation with "feedback as positive/negative experience" $r(90) = -.493$, $p < .01$.

- "Feedback easy to understand" showed a statistically significant correlation with "confidence that interpretation of results was correct", $r(90) = .297$, $p < .01$.
- "Confidence that interpretation of results was correct" showed a statistically significant correlation with "feedback as positive/negative experience", $r(90) = -.402$, $p < .01$.

Reliability

Cronbach's alpha was calculated for the interval-level questions within the categories Expectations regarding personality assessment, Experience of personality assessment, and Experience of receiving feedback (Appendix 21).

Three items were reverse scored for the purpose of calculating reliability ("expectation of personality assessment as positive/negative experience", "experience of personality assessment as positive/negative", and "feedback as positive/negative experience

The first set of items demonstrated low reliability (.245) but the second and third sets showed good reliability (.692 and .619 respectively).

5.1.5 Partial re-validation of SFSCS and RSES

As discussed in the previous chapter, because the SFSCS and RSES were not developed using New Zealand participants, their appropriateness cannot be assumed for use with a New Zealand participant group.

Whenever a psychometric tool is imported into a new context, its properties in that context should be re-examined. This includes checks of administrative viability, cultural appropriateness, and participant reactions, as well as revalidation in terms of reliability, and content-, criterion- and construct-related validity (Arnold & Smith, 2013; Cronbach & Drenth, 1973; Dana, 1993).

The following attempts were made to validate the use of these instruments in the New Zealand context, however time constraints made a comprehensive analysis impossible.

Where possible, factor structure has been compared to original data, and descriptive statistics, reliability, and convergent validity (correlating the two measures) have been calculated.

Both instruments were developed in North America where there are cultural overlaps with New Zealand culture. Both instruments were developed in English.

For the sake of completion of the project, and based on the following analyses, it was assumed that both instruments were suitable for use with New Zealand participants.

SFSCS

Descriptive statistics

Descriptive statistics for each subscale measured by the SFSCS were calculated for the participants in Sub-project B, before and after completing the personality assessment (Appendix 22).

Subscales were scored by summing response scores for items within each subscale. Power and Gifted subscales were weighted by .857 and 1.2 respectively, for the purpose of subscale comparison. A Total Self-concept score was derived by adding the subscales Likeability, Task Accomplishment, Power (weighted), Gifted (weighted) and Moral, and subtracting the subscale Vulnerability, as per Stake (1994).

Means for the SFSCS subscales varied before and after completion of the personality assessment, but retained the same ranking (in decreasing order): Moral, Task Accomplishment, Likeability, Gifted, Vulnerability and Power.

The Gifted subscale had the greatest range and the Vulnerability subscale had the largest standard deviation and the largest variance, both before and after completion of the personality assessment.

Task Accomplishment had the greatest absolute skew before completion of the personality assessment, and Likeability had the greatest absolute skew after completion of the personality assessment.

Task Accomplishment had the greatest absolute kurtosis before and after completion of the personality assessment.

Reliability

Cronbach's alpha was calculated for all subscales of the SFSCS as well as the total scale, for the participants in Sub-project B, before completing the personality assessment (Appendix 23). Vulnerability subscale item scores were reversed for the purpose of calculating reliability for the total scale.

The total scale demonstrated high reliability (α = .890), with two items ("Aggressive" and "Law-abiding") that would have improved the reliability of the scale if they had been removed.

All subscales (apart from Moral (α = .712)) demonstrated high reliability, with alpha coefficients ranging from .819 (Power) to .859 (Vulnerability). For Likeability, Power, Vulnerability, and Gifted there were no items that would have improved the reliability if they had been removed.

For the Task Accomplishment subscale one item ("Plans ahead") would have improved the reliability of the subscale if it had been removed. For the Moral subscale one item ("Law-abiding") would have improved the reliability of the subscale if it had been removed.

Validity

As in Sub-project A, a Principal Component Analysis (PCA) was attempted in order to locally re-validate for construct validity by checking the component/factor structure of the SFSCS in the New Zealand data (Bryant & Yarnold, 1995).

Specifically, a PCA was conducted on the 36 items of the SFSCS (for the participants in Sub-project B, before completing the personality assessment) with orthogonal rotation (VARIMAX). Multicollinnearity was a problem (correlation matrix determinant is 2.329E-10), so the PCA was recalculated using oblique rotation (direct oblimin; Field, 2009).

The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .774, and all KMO values for individual items were >.5. Bartlett's test of sphericity $\chi 2$ (630) = 1733.770, $p < .001$, indicated that correlations between items were sufficiently large for PCA.

An initial analysis was run to obtain eigenvalues for each component in the data. Nine components had eigenvalues over Kaiser's criterion of 1 and in combination explained 69.23% of the variance. However, the scree plot shows a clear point of inflection after six components – providing some support for the six factors extracted by Stake (Bryant & Yarnold, 1995; Appendix 24).

In validating the SFSCS, Stake (1994) theorised that the subscales of the SFSCS should be correlated more highly with a measure of self-esteem than with social desirability. Her results supported this theory, with a high correlation between the SFSCS and RSES (for Total Self-concept, $r = .62$, $p < .0001$).

Using Sub-project B data, a correlation matrix including the subscales of the SFSCS and the RSES was produced. Separate matrices were generated using data collected before and after completion of the personality assessment (Appendix 25).

Comparable to Stake's results, the correlation between Total Self-concept and Self-esteem (before completing the personality assessment) was $r(90) = .715$, $p < .01$; the median subscale correlation was .36. After completing the personality assessment, the correlation between Total Self-concept and self-esteem was $r(90) = .655$, $p < .01$, with a median subscale correlation of .34.

These correlations provide indirect support for the validity of the SFSCS with the participant group for Sub-project B, via a similar relationship between self-concept and self-esteem.

RSES

Descriptive statistics

Descriptive statistics for the RSES were calculated for the participants in Sub-project B, before and after completing the personality assessment (Appendix 26).

The RSES was scored by summing response scores for all items (with negative items scored negatively) to achieve a total Self-esteem score.

The mean showed a small increase after completion of the personality assessment and receipt of feedback, and the standard error decreased. The range, standard deviation, variance, skew and kurtosis were all lower after completion of the personality assessment, though absolute kurtosis increased.

Reliability

Cronbach's alpha was calculated for the RSES (before completing the personality assessment; Appendix 27).

The total scale demonstrated high reliability ($\alpha = .921$), and no items would have improved the reliability of the scale if they had been removed.

Validity

While a complete validation study for construct validity was beyond the scope of this thesis, an attempt was made to clarify whether New Zealand data would "reproduce" the component structure. Instead of exploratory factor analysis, Principal Component Analysis (PCA) was conducted on the 10 items of the RSES (for the participants in Sub-project B, before completing the personality assessment; Appendix 28). Orthogonal rotation (VARIMAX) was chosen (Bryant & Yarnold, 1995; Field, 2009).

The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .905. All KMO values for items were > .55. Bartlett's test of sphericity $\chi 2$ (45) = 585.945, $p < .001$, indicated that correlations between item variables were sufficiently large.

The nature of components as well as eigenvalues were reviewed. Two components had eigenvalues over Kaiser's criterion of 1.0. This combination explained 70.39% of the variance.

5.2 Sub-project B: Results

5.2.1 Results of testing

Following on from analyses of the psychometric properties of the data collection instruments, formal testing of the below hypotheses (including pre-test checks) was conducted.

Hypothesis 2 states that completion of a personality assessment and receipt of results/feedback will be associated with a change in:

- Any or all of the following components of self-concept as measured by the Six-Factor Self-Concept Scale: Likeability, Task Accomplishment, Power, Vulnerability, Gifted, Moral; and/or,
- Total self-concept as measured by the Six-Factor Self-Concept Scale; and/or,
- Self-esteem as measured by the Rosenberg Self-Esteem Scale

Hypothesis 2a states that positive or negative change to self-concept and/or components of self-concept and/or self-esteem will be influenced by any of the following variables:

- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results
- Sex
- Ethnicity
- Age

- Previous experience of personality assessment(s)
- Reason(s) for prior completion of personality assessment(s)

Calculation of change scores

A "change score" for each self-perception variable was calculated by subtracting the first-wave value of self-concept/self-esteem (before personality assessment, Questionnaire 1) from the second-wave (Questionnaire 2). The use of such change scores was addressed in Chapter 4.

Detail regarding the theoretical basis, advantages, and limitations of change scores will be revisited in the Discussion chapter.

Analysis of distributions

As discussed above, variables were checked for the properties of their distributions (Appendices 22 and 26). For each variable, indices of central tendency (mean, median, mode), spread (variance and SD) as well as skewness and kurtosis statistics were generated.

The dependent variables in relation to self-perception had distributions with skewness ranging from -1.106 to .166 before completion of the personality assessment. The same indicators ranged from -1.103 to .316 after completion of assessment. Kurtosis ranged from -.532 to 1.444 before completion of personality assessment; from -.532 to 1.444 after completion of assessment. Such degrees of skewness and kurtosis are commonly estimated for psychometric score variables with a small N and non-representative samples.

Distributions of the change scores were also examined. The variables of Self-esteem and the Likeability subscale were more skewed than the other change score variables (Appendix 29). Power, Moral and Self-esteem also showed significant kurtosis.

Kolmogorov-Smirnov tests were then performed to see the extent to which these variables deviate from symmetric distributions (Appendix 30). Among the change score variables, the distribution for Total Self-concept was closest to symmetric.

Change in self-perception over time

To test Hypothesis 2, a paired-samples t-test was conducted to compare self-concept subscales, Total Self-concept and Self-esteem scores before and after completion of the personality assessment (Appendix 31).

Of the eleven t-values generated, most fell under 1.5 and did not reach a level of significance, suggesting minimal temporal shift or the absence of shift. Two comparisons showed a significant effect: the Likeability component of self-concept, and Self-esteem.

On average, participants' Likeability scores after completing the personality assessment (M = 33.05, SE = .60) were higher than before completing the personality assessment (M = 32.45, SE = .56). This difference, -.609, BCa 95% CI [-1.212, -.006], was significant $t(91) = -2.005$, $p = .048$, and represented a small-sized effect, $r = .21$.

On average, participants' Self-esteem scores after completing the personality assessment (M = 21.13, SE = .59) were higher than before completing the personality assessment (M = 20.60, SE = .61). This difference, -.533, BCa 95% CI [-1.065, .000], was significant $t(91) = -1.988$, $p = .050$, and represented a small-sized effect, $r = .20$.

Using a Bonferroni correction (Bonferroni, 1935, 1936) for repeated t-tests, $p < .05$ was replaced by the more conservative requirement $p < .006$. At this level, both the effect for Likeability and the effect for Self-esteem are no longer significant.

The direction of potential shift, i.e. increase versus decrease of score after completion of the personality assessment, was also of interest. While not all t values produced were significant, demonstrated tendencies appear to be parallel. All subscales of self-concept (other than Vulnerability and Moral), as well as Total self-concept showed an increase in means following completion of the personality assessment.

Variables influencing change in self-perception

To partially test Hypothesis 2a, correlations were calculated between the variables below and each of the change score variables in order to assess whether these interval-level variables may have influenced any change to self-perception.

- Expectation of personality assessment as positive/negative experience
- Confidence in correct interpretation of personality assessment results

There was a minor but significant correlation (.209, $p < .05$) between "expectation of personality assessment as positive/negative experience" and the dependent variable Change in Task Accomplishment (Appendix 32).

There were no other significant correlations between either of the above variables and the dependent variables.

For a multivariate approach to testing Hypothesis 2a, several analyses of variance were conducted via MANOVA. In the output, Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root were evaluated (Field, 2009).

Change score variables for self-perception served as dependent variables. Specifically, the self-concept subscales of SFSCS (Likeability, Task Accomplishment, Power, Vulnerability, Moral, Gifted), Total Self-concept (measured by the SFSCS), and Self-esteem (measured by the RSES) were the dependent variables.

Independent variables were:

- MANOVA 1 independent variables: sex; ethnicity; age
- MANOVA 2 independent variables: "Previous experience of personality assessment(s)", "reason(s) for prior completion of personality assessment(s)"

Using Roy's Largest Root, MANOVA 1 demonstrated a significant effect of age on the dependent variables, $\Lambda = .232$, $F(7,73) = 2.42$, $p < .05$. There was also a significant effect of age + sex on the dependent variables, $\Lambda = .246$, $F(7,71) = 1.42$, $p < .05$), and a significant effect of age + ethnicity on the dependent variables, $\Lambda = .227$, $F(7,71) = 2.30$, $p < .05$. However, separate ANOVAs on the independent variables revealed no significant F value representing any effects on the dependent variables (Appendix 33).

MANOVA 2 did not demonstrate any significant effects of "previous experience of personality assessment(s)" and/or "reason(s) for prior completion of personality assessment(s)" on the dependent variables. Separate ANOVAs on the independent variables revealed no significant F value representing any effects on the dependent variables.

5.2.2 Intercorrelation of self-perception variables

The variables self-concept and self-esteem were expected to be associated with each other. In fact, as discussed above, correlation between the SFSCS and the RSES was counted as evidence for the convergent validity of the SFSCS (Stake, 1994; Appendix 25).

There were significant correlations between Self-esteem and all subscales of self-concept within the data collected before completion of the personality assessment: Likeability (.250), Task Accomplishment (.462), Power (.454), Vulnerability (-.573), Gifted (.571) and Moral (.263).

In the data collected after completion of the assessment, Self-esteem showed a significant correlation with all subscales of self-concept: Likeability (.334), Task Accomplishment (.448), Power (.343), Vulnerability (-.537), Gifted (.448) and Moral (.314).

Change in Self-esteem correlated significantly with Change in Vulnerability (-.239) and Change in Gifted (.250).

Change in Likeability correlated significantly with Change in Task Accomplishment (.402), Change in Gifted (.329), and Change in Moral (.249). Change in Task Accomplishment also correlated significantly with Change in Power (.253), Change in Gifted (.374) and Change in Moral (.328). Change in Power also correlated significantly with Change in Gifted (.316).

See Appendix 34 for correlations between change scores for the SFSCS and RSES.

6 Discussion

Sub-projects A and B sought to explore whether completion of a personality assessment (or completion of a personality assessment and receipt of written feedback) was associated with a change in self-perception. While the results show preliminary support for the hypothesis that such an association exists, the outcomes of the two sub-projects signal trends in opposing directions, and implicate different elements of self-perception. Moreover, the use of a Bonferroni correction, which is justified in the methodological setting, relativises or removes the evidence of temporal shift in Sub-project B.

This chapter will discuss how the results from each sub-project relate to each other and to the extant literature. It will also examine the limitations of the sub-projects within this thesis, and outline considerations for future research.

6.1 Outcomes of hypothesis testing

6.1.1 Hypotheses for Sub-project A

- Hypothesis 1: Completion of a personality assessment as part of a selection process will be associated with a change in self-concept (including components thereof), and/or self-esteem.
- Hypothesis 1a: Any/all of the following variables may influence this change: Number of roles applied for in last month, expectation of personality assessment as positive/negative experience, confidence in correct interpretation of personality assessment results, confidence in correct interpretation of ability assessment results, sex, ethnicity, age, current employment status, reason for applying for role, expected salary, previous experience of personality assessment(s), and purpose of previous experience.

Hypothesis 1 was supported with regard to Power (weighted), a component of self-concept. On average, participants' weighted Power scores were significantly lower after completing the personality assessment (M = 23.02, SE = .97) than before completing the personality assessment (M = 24.54, SE = .90), $t(26) = 3.34$, $p = .003$, $r = .30$.

Neither correlational analyses nor analyses of variance (ANOVA, MANOVA) provided support for Hypothesis 1a.

6.1.2 Hypotheses for Sub-project B

- Hypothesis 2: Completion of a personality assessment and receipt of results/feedback will be associated with a change in self-concept (including components thereof), or self-esteem.
- Hypothesis 2a : Any/all of the following variables may influence this change: Expectation of assessment as positive/negative experience, confidence in correct interpretation of personality assessment results, sex, ethnicity, age, previous experience of personality assessment(s), purpose of previous experience.

Hypothesis 2 was weakly supported with regard to Likeability, a component of self-concept, and Self-esteem. On average, participants' Likeability scores were higher after completing the personality assessment (M = 33.05, SE = .60) than before completing the personality assessment (M = 32.45, SE = .56), $t(91) = -2.005$, $p = .048$, $r = .21$; and on average, participants' Self-esteem scores were higher after completing the personality assessment (M = 21.13, SE = .59) than before completing the personality assessment (M = 20.60, SE = .61, $t(91) = -1.988$, $p = .050$, $r = .20$.

Both of these t-test outcomes no longer satisfied the probability for significance after a Bonferroni correction recognising serialised tests. For a contemporary reconsideration of Bonferroni's solution, see Abdi (2007) and Perneger (1998).

Neither correlational studies nor analysis of variance (MANOVA) provided robust support for the self-perception relationships involving expectation of assessment as positive/negative experience, or confidence in correct interpretation of personality assessment results. Using change scores of self-concept and self-esteem as dependent variables, it appears that subjective experience and expectations, as assessed by researcher-created questionnaire items, did not have the association or impact queried by Hypothesis 2a.

An exception to this pattern is the correlation between reported "expectation of personality assessment as positive/negative experience" and Task Accomplishment (a component of self-concept; using change score of Task Accomplishment). A positive correlation coefficient signals that the more favourable the participant expects the personality testing experience to be, the more this particular aspect of self-concept will "gain". However, with an r barely over .2 (coefficient of determination around .04), the independent variable (experience expectation) only accounts for a minimal proportion of the variance of the dependent variable (one component of self-concept). The fact that other components of self-concept measured

simultaneously by the same self-concept test did not produce a significant correlation also warns that the Task Accomplishment result may be weak and isolated, rather than warranting a theoretical interpretation.

6.2 Integrating results for Sub-Projects A and B

6.2.1 Comparing Sub-projects A and B

While the two sub-projects were similar, there were differences in design and procedure.

Sub-project A measured self-perception before and after completion of the personality assessment alone, and no participant received feedback from their personality assessment. Therefore any consequences of receiving feedback could not be measured for this participant group.

Sub-project B measured self-perception before the personality assessment, and after both completion of the personality assessment and receipt of feedback. Therefore any consequences of the personality assessment alone could not be separated from any consequences of receiving feedback.

The context of completion of the personality assessment was also different for participants within each sub-project.

In Sub-project A, participants were job applicants and the personality assessment formed part of testing entailed by a genuine job application. Outcomes of the personality assessment had real-world consequences in terms of success or failure (at being offered the job for which they had applied). It is likely that the experience of completing the personality assessment would have been (accurately) linked with increased/decreased likelihood of success on the basis of the personality assessment results. Participants may have had reactions to the perceived relevance/fairness of the use of the personality assessment in the selection process (Gilliland, 1993).

Participants in Sub-project A also completed and reflected on the personality assessment knowing that their results would be viewed by the members of the selection panel (as well as the recruitment coordinator) at the hiring organisation.

The participants in Sub-project B were students who volunteered to complete the personality assessment for this research. There were no real-world consequences related to their assessment results, and they were guaranteed the opportunity to receive written feedback.

Therefore even if negative results were anticipated at any stage of the process, these are more likely to have been viewed as a constructive source of information with which to strengthen future job applications/ personality assessment completions.

Unlike participants in Sub-project A, those in Sub-project B knew that the results of their personality assessment would not be made available to anyone but the researcher. Therefore any potential negative expectations regarding the results were unlikely to trigger fear of judgement or embarrassment (emotions which may have informed the experience of participants in Sub-project A).

6.2.2 Interpretation of differences in results for Sub-projects A and B

The design and procedural differences between Sub-projects A and B mean the results for each sub-project are not directly comparable.

Hypothesis testing for Sub-project A demonstrated a relationship between completion of the personality assessment and change in self-perception.

It is assumed that either:

- The results for hypothesis testing in Sub-project B contain the same effect on self-perception of the personality assessment (perhaps as well as an additional effect related to the written feedback); or,

- There was a sufficient difference between the two sub-projects such that the results for Sub-project B reflect a different relationship between completion of the personality assessment and change to self-perception (with or without the addition of an effect related to the written feedback).

Assuming that the reported results are not due solely to error, possible interpretations are described below.

Feedback as the only point of difference

One explanation for the results of both sub-projects is that a) both participants groups were affected in the same way by the personality assessment, but b) receiving the written feedback (Sub-project B) exerted an additional effect.

That is, both participant groups may have experienced a decrease to Power, as a component of self-concept, following completion of the personality assessment, but the written feedback provided to participants in Sub-project B had the subsequent effect of increasing Power (such that no overall decrease was observed) and also increasing Likeability, as a component of self-concept, and total Self-esteem (resulting in an observed increase).

This speculation rests on a fairly tentative conjecture that - if measurement had been viable immediately after completion of the personality assessment - participants in Sub-project B would have demonstrated a similar decrease in Power and no increase in Likeability or Self-esteem. In addition, one needs to assume that participants in Sub-project A would have demonstrated no significant change in Power and a significant increase in Likeability and Self-esteem if they had received the written feedback that was provided to Sub-project B participants, prior to the second measurement of self-perception.

Context of personality assessment as the point of difference

Another possibility is that the difference in context for participants in Sub-projects A and B led to a different impact on self-perception as a result of completing the personality assessment. This could occur via the context producing a different experience of completing the personality assessment itself, leading to a different effect on self-perception.

For example, the difference in demand characteristics of the personality assessment in relation to securing a job offer versus gaining constructive feedback is likely to have resulted in a different experience of completion of the personality assessment.

Participants may also have had different motives for participation in the relevant sub-project, or the varied contexts of completion may have created contrasting psychological settings from which participants reflected on their experience of the personality assessment.

Participant characteristics as the point of difference

Still another interpretation is that the participants themselves were dissimilar in some significant way across Sub-projects A and B.

For example, the participants in Sub-project A were job applicants applying for roles at one specific organisation. These participants might have a unique perspective on, or reaction to, the completion of personality assessments.

Alternatively, it may be that students, who are regularly tested for diverse purposes, have a unique response to tests such as personality assessments (Kam, Wilking, & Zechmeister, 2007).

In support of the second and third interpretations, a number of researcher-created items showed different response distributions in Sub-projects A versus B in relation to experience of the personality assessment. For example, a much greater proportion of participants in Sub-project B (48%) said that completion of the personality assessment was a (somewhat or very) positive experience than did participants in Sub-project A (22%).

6.3 Relating New Zealand findings to published models and results

The goal of this section is to analyse the findings for Sub-projects A and B in order to link them to pre-existing models and theories, as well as published evidence in support of them. Possible reasons for apparent departures from global findings will be considered.

As discussed previously, little research could be found regarding the consequences of personality assessment for selection, with contributions from Industrial/Organisational (I/O) Psychology limited to 'applicant reactions' research. This area has yet to investigate applicant reactions in the form of changes to any self-referent construct as a result of personality assessment. (Less direct links between the current findings and published applicant reactions research are discussed further below).

In light of this, examination of the literature was broadened to include traditions outside of I/O Psychology. Research methodologies other than real-life observational research were also considered, in relation to antecedents of change in self-perception as well as consequences of personality assessment.

6.3.1 Effects of positive and negative events

As discussed in the Literature Review chapter, Stake, Huff and Zand (1995) conducted two studies in which students were exposed to guided imagery scenarios presenting either positive or negative self-relevant information, such as social acceptance or unexpectedly scoring poorly on an important test.

After each scenario, participants provided a rating to indicate the extent to which they anticipated that the scenario would enhance or lessen their view of themselves, as well as ratings to indicate the extent to which they anticipated that the scenario would affect their

view of themselves in specific relation to each of the six components of self-concept measured by the SFSCS. Participants also completed the RSES measure of self-esteem (used in Sub-projects A and B as the measure of self-esteem), prior to the experiment.

Stake et al. found that anticipated self-perception was impacted by the positive and negative events, with greater shifts in self-perception occurring as a function of pre-existing self-esteem. Participants with low self-esteem were more likely to anticipate greater reductions in self-concept (and greater negative generalisation across components of self-concept) as a result of the negative events.

Sub-projects A and B did not replicate these findings.

Participants in Sub-project A were asked directly whether completing the personality assessment was a positive or negative experience, but their responses were not significantly correlated with any self-concept subscale scores, nor Total Self-concept or Self-esteem after completion of the personality assessment.

Participants in Sub-project B were also asked this question, and their responses were significantly correlated (one-tailed) with the Likeability component of self-concept only, after completion of the personality assessment and receipt of feedback.

Participants in Sub-project B were also asked whether receiving the written feedback was a positive or negative experience, and their responses to this item correlated with the Likeability, Task Accomplishment, and Moral components of self-concept, as well as Total Self-concept after completion of the personality assessment and receipt of feedback.

Greater associations were revealed when the results were split by low, medium or high Self-esteem scores, prior to completion of the personality assessment (a categorisation suggested by Stake et al., 1995).

In Sub-project A, the low Self-esteem group showed correlations between positive/negative experience and self-concept and Self-esteem after completion of the personality assessment, however there were only two participants in this sub-group so these results are not interpretable. The medium Self-esteem group (n = 18) showed a significant correlation between positive/negative experience and the Likeability component of self-concept as well as Self-esteem after completion of the personality assessment. The high Self-esteem group (n = 7) showed a significant correlation between positive/negative experience and the Vulnerability component of self-concept after completion of the personality assessment.

In Sub-project B, the low Self-esteem group (n = 15) showed a significant correlation between positive/negative experience and the Task Accomplishment component of self-concept after completion of the personality assessment and receipt of feedback. The medium Self-esteem group (n = 58) showed a significant correlation between positive/negative experience and the Power component of self-concept after completion of the personality assessment and receipt of feedback. The high Self-esteem group (n = 19) showed a significant correlation between positive/negative experience and the Likeability, Task Accomplishment and Moral components of self-concept after completion of the personality assessment and receipt of feedback.

Also in Sub-project B, the low Self-esteem group (n = 15) showed a significant correlation between 'Was receiving the written feedback a positive or negative experience?' and the Task Accomplishment and Power components of self-concept after completion of the personality assessment and receipt of feedback. The medium Self-esteem group (n = 58) showed a significant correlation between this item and the Likeability and Vulnerability components of self-concept after completion of the personality assessment and receipt of feedback. The high Self-esteem group (n = 19) showed a significant correlation between this item and the Likeability and Task Accomplishment components of self-concept as well as Total Self-concept after completion of the personality assessment and receipt of feedback.

The differences between the current findings and the Stake et al. results are at least partially explained by methodological features.

a) Participants: Participants in the Stake et al. studies were all university students; participants in Sub-project A were job applicants.
b) Genuine event: While the purpose of the Stake et al. research was to gauge the impact of real-life events, the studies used manipulation customary in experimental social psychology to simulate such events in a laboratory. That is, while events inclusive of positive or negative self-relevant information were portrayed via guided imagery, no real-life self-relevant information outside the laboratory was received by any participant.
   In contrast, findings for Sub-project A were based on data from job applicants completing a personality assessment as part of a genuine selection process. Even in Sub-project B, the feedback received after completion of the personality assessment was real-life feedback from the assessment provider after actual assessment, though participants were university students rather than job applicants.
c) Measured shift versus anticipation: The dependent variable in the Stake et al. studies was the participants' anticipated reactions, rather than changes across successive measures of self-perception, as in Sub-projects A and B.

d) Cumulation: While Stake et al. state that they adopt the theoretical model underlying the SFSCS, they did not use this instrument in full, nor refer to specific findings in relation to the separate components. Rather, the components of self-concept were used to indicate 'generalization' of impact across self-concept only.

Considering these four concrete dissimilarities, the use of genuine events and measured shifts in self-perception (in Sub-projects A and B) appears to provide the most likely explanation of why the current findings contain less uniform trends across the self-perception variables in terms of the consequences of positive and negative experience.

6.3.2 Alternative theoretical frameworks

Effects of self-relevant information

Beyond Stake, Huff and Zand (1995), there are few studies that are able to directly contribute to interpretation of the findings for Sub-projects A and B.

However, in the broader literature within which the Stake et al. research is situated, there is a significant body of work regarding the effects of self-relevant information on self-perception (vanDellen, Campbell, Hoyle, & Bradfield, 2010), which can provide potentially meaningful, albeit less explicit, comparisons.

Self-relevant information (or 'feedback to the self-system') can come from the reactions or behaviour of others (such as in (Stake et al., 1995), or perception of one's own behaviour (Markus & Wurf, 1987). Both sources of information can also be mutually informed by one another (Fazio et al., 1981).

As discussed in the Literature Review chapter, a range of social psychology experiments have demonstrated that manipulating subjects' behaviour (or self-presentations) can result in a change to self-perception in the direction of the manipulated behaviour/presentation (see Fazio et al., 1981; E. E. Jones et al., 1981; Markus & Kunda, 1986; Rhodewalt & Agustsdottir, 1986; Schlenker et al., 1994; Tice, 1992).

For example, as cited earlier, participants who were induced to portray themselves as emotionally stable (or emotionally responsive) showed a subsequent change in self-reported self-concept to be more emotionally stable (Tice, 1992).

Linking the current findings to the above models requires that we treat responses to personality assessment items as self-presentations, possibly influenced by demand

characteristics of the selection context. Johnson (1981) discusses the contrasting views of self-report and self-presentation and provides empirical support for personality assessment responses as self-presentation. This view also aligns with Hogan's Socioanalytic Theory of personality (R. Hogan & Foster, 2016).

In Sub-project A, participants' mean scores on the Power component of self-concept decreased after completion of the personality assessment. The Power subscale is described as representing "qualities of strength, leadership and persuasiveness" (Stake, 1994, p. 69).

Of the 27 participants in Sub-project A, 21 were applying for a low level role. It is conceivable that these applicants realised that the role they were applying for did not require "Power" characteristics, and downplayed these qualities in their responses within the personality assessment. Prien, Schippman and Prien (2003) showed that applicant numerical ability can be downplayed in applications for roles where high numerical ability is not required as compared to test scores outside of selection.

If demand characteristics induced participants in Sub-project A to answer in a way that decreased self-reported levels of strength, leadership and persuasiveness, then it would follow that the subsequent measure of self-concept would show a decrease in Power.

In Sub-project B, participants' mean scores increased on the Likeability component of self-concept, and on Self-esteem, after completion of the personality assessment and receipt of written feedback. The Likeability subscale is described as, "qualities of interpersonal accessibility" (Stake, 1994, p. 69), and Self-esteem entails "the individual's positive or negative attitude toward the self" (Rosenberg et al., 1995, p. 141).

In Sub-project B there were no motivating consequences linked to the personality assessment results. It is therefore unlikely that participants would have perceived significant demand characteristics (with a minor possibility that participants could have viewed their Likeability scores as affecting their chance of winning the prize-draw, even though this was presented as a random draw).

For Sub-project B, it is assumed that the written feedback (rather than the personality assessment) is the more likely cause of shift in self-perception, presumably by triggering biased scanning (see E. E. Jones et al., 1981). That is, the positively worded feedback statements accentuated participants' perceived qualities related to the Likeability component of self-concept and Self-esteem, leading to an increase in scores of these constructs.

Response construction during personality assessment

There are cognitive processes underlying completion of personality assessments, though these are common across all self-report questionnaires, i.e. they are not specific to personality assessment completion.

As discussed in the Literature Review, Feldman and Lynch (Feldman, 1992; 1988) posit that where a response to a question does not exist in memory, a response is constructed from available information. The theory also asserts that response construction occurs when a previously constructed response is not easily accessible in memory. Such 'constructive' processes may enhance understanding of the current findings.

Participants in both Sub-projects A and B were asked "Were there any questions in the personality assessment that asked you something about yourself that you had never considered before?" Results in relation to this item reveal that response construction may have occurred during completion of the personality assessment.

In Sub-project A, five of 27 participants responded Yes to this item; in Sub-project B 22 of 92 participants responded Yes.

It is assumed that at least for these subsets of participants, response construction, in the sense of Feldman, was a prevalent mechanism during completion of the personality assessment. It is also likely that response construction occurred for many additional participants facing questions about personality that they had previously never considered, who had no ready-made response retrievable from memory.

Of the 22 participants in Sub-project B who affirmed that the personality assessment had asked them questions regarding themselves that they had never considered before, six participants agreed that the personality assessment caused them to think differently about themselves (as well as another 16 participants who responded negatively to the first item). It is notionally possible that the mechanism involved in 'thinking differently about oneself' is response construction, and that for participants who responded Yes to the second item but not the first, this is due to response construction on the basis of information that was inaccessible rather than non-existent.

6.3.3 Applicant reactions to selection processes

As noted earlier, applicant reactions research has few studies examining the consequences for self-perception of selection processes or the consequences of personality assessment within

selection, let alone the consequences for self-perception of personality assessment within selection.

The organisational justice model (Gilliland, 1993) proposes two antecedents to outcomes related to self-perception (included in this category are self-esteem, self-efficacy and future job-search intentions). These are: overall fairness of selection procedures, and overall fairness of selection outcomes. (See also Van Vianen, Taris, Scholten, & Schinkel, 2004).

As the current findings included only one participant in a selection process who had been notified of the selection outcome, the latter antecedent could not be examined in relation to Sub-project A or B. However, with regard to 'Overall fairness of selection process', the following variables from the current research are pertinent to procedural rules from the organisational justice model. These were treated as indicators of perceptions of fairness, and were correlated with changes to self-perception for participants in Sub-project A only.

Table 11: Procedural rule indicators

Procedural Rule	**Sub-project A questionnaire item**
Job relatedness	Do you think the personality assessment was relevant to your job application?
	Were there any questions that did not apply to you as a person?
Opportunity to perform	Now that you have completed the personality assessment, select from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results.
	Did the personality assessment allow you to express a true and complete picture of your personality?
	Did you have any difficulty completing the personality assessment?

Twenty-six percent of participants said they thought that the personality assessment was not relevant to their job application. This was significantly correlated with the Vulnerability component of self-concept as well as Total Self-concept after completion of the personality assessment.

Thirty-seven percent of participants said that they had difficulty completing the personality assessment. This item was significantly correlated with the Likeability component of self-concept after completion of the personality assessment, and also had a statistically significant correlation with the item, “Did completing the personality assessment have any impact on your optimism about successfully gaining the job you applied for, or future jobs?”

In response to the question, “Now that you have completed the personality assessment, select from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results” 4% responded with 0% confidence, 15% responded with 25% confidence, 44% responded with 50% confidence, 33% responded with 75% confidence and 4% responded with 100% confidence. Scores on this item were significantly correlated with the item, “Did the personality assessment allow you to express a true and complete picture of your personality?”

Thirty-seven percent of participants responded Yes to the question, “Were there any questions that did not apply to you as a person?” This item was significantly correlated with the Task Accomplishment component of self-concept, after completion of the personality assessment.

These items provide some support for the proposed relationship between applicant perceptions of fairness and self-perception.

6.4 Limitations of the current research

6.4.1 Adjustments to original research design

Longitudinal design

Having to simplify the original research design imposed a major limitation on studying the time-course of any change in self-perception. The originally proposed design of the research was longitudinal in its true sense. It included a third wave of measurement intended to record changes after applicants had received feedback about their personality assessment performance.

Logistics, particularly participant recruitment and retention, necessitated that a longitudinal data collection had to be abandoned. Neither sub-project was therefore able to probe into stability over time of changes to self-perception, in the customary sense of stability. One of the major advantages of the longitudinal design was lost, severely limiting the type of results that were generated and the interpretations that were available.

Control group

The original design for Sub-project A also included a control group - employees going through a similar selection process who were not required to complete a personality assessment. Because Sub-project A sourced job applicants from one organisation only (and that organisation used personality assessment in all selection processes), there was no opportunity to gain participants who were going through a selection process without personality assessment. The lack of a control group indicates a major shortcoming; i.e. there is no justification to state with confidence that changes to self-perception are due to completion of the personality assessment (and/or receipt of feedback).

6.4.2 Local use of imported psychometric instruments

Adaptation

Due to the logistics and resources of this thesis, a proper psychometric adaptation of the SFSCS and RSES in the New Zealand context was not feasible. The Results sections summarise partial attempts to obtain validity estimates for assessments, with mixed results.

It is common to use British and American tests in New Zealand with minimal or zero adaptation; however, the ubiquity of such practice does not justify erroneous assumptions, nor obviate professional norms. Psychometric tests used in a new environment after being imported from elsewhere, are required to undergo careful re-examination and adaptation (Cronbach & Drenth, 1973). A professional stipulation for adapting tests (Cuellar & Paniagua, 2000) is that an integral research project checking viability of administration, test-taker reactions, cultural appropriateness, reliability estimates, and validity levels must be conducted in the new context. Revalidation in terms of content-, criterion-, and construct-related validity is paramount (Dana, 1993). It is not merely the use of a test in a different country such as New Zealand, but utilisation in particular local contexts, test-taker groups, and organisations that prompts concerns about reliability and validity (Samuda, 1998; Suzuki, Meller, & Ponterotto, 1996).

While the SFSCS and RSES are English-language tests in which items are readily understood by the participants in this project, and there are assumed cultural overlaps between North America and New Zealand, responsible psychometric assessment cannot take for granted that validity and reliability levels are “transported” (Dana, 1993). Ideally, a pilot project would have been conducted to examine the properties and functionality of both SFSCS and RSES with the relevant New Zealand target groups. Full construct validation regarding all sub-constructs of

self-concept and self-esteem would have to greatly improve upon the attempts reported in the Results sections, preferably employing convergent and discriminant validation, or the Multitrait-Multimethod Matrix (D. T. Campbell & Fiske, 1959).

Sources of measurement error

Participants completed the questionnaires online, therefore the situational variables of their settings were neither controlled nor recorded within this research. It is also possible that some participants did not complete all/any of the online questionnaires or personality assessments themselves. These factors may have introduced measurement error to the data. For discussion regarding measurement error caused by the use of 'change scores', see section 6.4.4 below.

There are also potential sources of measurement error in the form of order effects, as a result of the repeated measures research design (Wilson & Joye, 2017).

There are four order effects that could have affected participant results.

- Practice effects could have allowed participants to 'improve' their performance on the self-perception measures. Given that the measures are not testing ability, i.e. there are no right or wrong answers; improvement per se is unlikely. However, participants may respond differently to questions of self-concept and self-esteem the first time they are exposed to them versus later.
- Fatigue effects are more likely in the current studies, as participants may have grown tired of the lengthy questionnaires by the time they were completing the second measures of self-perception, and responded less reliably for this reason.
- History effects could have come into play during the time period between the first and second questionnaires. Due to the research occurring remotely via online questionnaires, there is no way of knowing what history effects may have occurred.
- Maturation effects refer to changes in participants due to the passage of time rather than the independent variable. The stability of self-perception is a direct concern in this research, and as with history effects, cannot be estimated.

6.4.3 Issues of sampling and generalisation

The goal of almost all empirical research is to discover results that apply to an entire population. This is typically done by collecting data from a subset of the population (sample) and generalising the findings to the rest of that population (Field, 2009; Newmark, 1997).

Sampling

Sampling theory outlines the requirements for the optimisation of generalisability by gaining a representative sample: the effective management of population definition, sample design, and sample size (Frankfort-Nachmias, 1996).

Population definition

The population of interest for this thesis was the New Zealand Labour Force. These are the people who are likely to be directly affected by personality assessment for selection by way of applying for a job at an organisation that uses personality assessment as part of its selection process.

The New Zealand Labour Force is comprised of everyone who is part of the usually resident, non-institutionalised, civilian population of New Zealand aged 15 and over, and who is either employed, or unemployed and looking for work (Statistics New Zealand, 2017). However, this population definition could not guide the current project in a strict sense, as it was practically impossible to obtain a listing of all persons included.

Sample design

Because a sampling frame (or list of every sampling unit/member of the population) was not available for the population specified above, probability sampling was not possible. For practical reasons including budget and time constraints, a convenience sample (also called an ad hoc sample (Frankfort-Nachmias, 1996)) was used.

As a non-probability sample design (Hansen et al., 1993), convenience samples offer no assurance that every sampling unit has a chance of inclusion in the sample (Frankfort-Nachmias, 1996). Because of this, convenience samples also offer no way of ascertaining the representativeness of the sample and therefore population parameters cannot be estimated on the basis of sample statistics (Frankfort-Nachmias, 1996).

Kline (2013) discusses mitigation of this issue via a posteriori documentation of specific sample characteristics for comparison with known population parameters. On this basis, participants in Sub-projects A and B were compared with the original population of interest (the New Zealand Labour Force) with regard to sex, age group, and ethnicity.

According to these variables, neither participant group was representative of the New Zealand Labour Force (Statistics New Zealand, 2017).

- Both participant groups had a higher proportion of females to males, while the Labour Force had a higher proportion of males to females.
- With regard to age group, the participants in Sub-project A over-represented 25-34 year olds, and under-represented all other age groups, while the participants in Sub-project B over-represented both 15-24 year olds and 25-34 years olds, and under-represented those 45 years and above.
- With regard to ethnicity, both participant groups were comparable to the Labour Force in that the clear majority of people identified themselves as European, followed by identification as Asian or Maori. However, both participant groups over-represented Europeans and under-represented Maori, and had no Pacific Peoples (or Middle Eastern/Latin American).

Table 12: Comparison of participant groups' Sex with New Zealand Labour Force

Sex	**Sub-project A**	**Sub-project B**	**NZ Labour Force June 2017**
Females	67%	89%	47%
Males	33%	11%	53%

Table 13: Comparison of participant groups' Age group with New Zealand Labour Force

Age group	**Sub-project A**	**Sub-project B**	**NZ Labour Force June 2017**
15 to 24 years[17]	11%	33%	15%
25 to 34 years	63%	32%	21%
35 to 44 years	7%	20%	19%
45 to 54 years	11%	13%	21%

[17] Statistics New Zealand Labour Force data begins at age 15 (Statistics New Zealand, 2017), whereas Studies 1 and 2 excluded participants under 16 years of age.

55 to 64 years	7%	2%	17%
65 years onwards	0%	1%	6%

Table 14: Comparison of participant groups' Ethnicity with New Zealand Labour Force

Ethnicity - Statistics NZ Level 1 Classification	**Sub-project A[18]**	**Sub-project B[18]**	**NZ Labour Force June 2017**
European	81%	87%	75%
Maori	7%	8%	12%
Pacific Peoples	0%	0%	6%
Asian	11%	5%	14%
Middle Eastern/Latin American	0%	0%	1%

Generalisation

The issues of sampling identified above clearly demonstrate that the results of Sub-projects A and B can be considered local at best, and that generalisation from them is not justified.

However, this does not necessarily imply lower worth of findings (Wilson & Joye, 2017), as they can provide a starting point from which to launch more methodologically rigorous investigations in this area. Numerous published studied in Organisational Psychology use non-samples and preclude generalisation. Additionally, while the sampling issues are sufficient to render the results non-generalisable, there may be other aspects relevant to considering possible generalisation.

Generalisability is closely allied with external validity (Frankfort-Nachmias, 1996; Kline, 2013). Kam et al., (2007) discuss three aspects to external validity: instrumentation, subjects compared to others within the population, and subjects compared to other populations of interest.

[18] Statistics New Zealand Labour Force data allows recording of multiple ethnicities, whereas Studies 1 and 2 allowed participants to select only one ethnicity (Statistics New Zealand, 2017).

- 'Instrumentation' refers to treatments and measurement, e.g. how closely the selection process experienced by participants in Sub-project A resembles other selection processes, and how similar the experience of completing the OPQ-32R is to completing other personality assessments.
- 'Subjects compared to others within the population' refers to representative sampling as described above, and so will not be discussed any further in this section. It also considers whether there are qualitative differences between those who self-select to take part in research versus those who are invited but choose not to participate.
- 'Subjects compared to other populations of interest' refers to how closely the population of interest resembles other potential populations of interest, e.g. job applicants in Australia.

The original design for Sub-project A hoped to include participants involved in other selection processes (i.e. at other organisations), using personality assessments other than the OPQ-32R, in order to compare the effects of different tests. With only one personality assessment used in both sub-projects, it is not possible to state with confidence that the results found here are not restricted to the OPQ-32R.

The original design also hoped to include measurement of employees receiving personality assessment results/feedback from different employer representatives. In real life, feedback following a personality assessment is not typically delivered in written format, but via a face-to-face meeting or phone conversation. Feedback delivered in this way could have a different impact to the written feedback received by participants in Sub-project B.

6.4.4 The use of change scores

The objective of the current project invited the use of 'change scores' to estimate change over time (Collins, 1996). A 'change score' (also termed 'gain score', 'raw score', 'obtained gain', 'raw change score', 'difference score' or 'shift score') is interpreted in this context as the numerical difference between scores produced at times t_1 and t_2, measuring the same construct via the same test (Cronbach & Furby, 1970; Humphreys, 1996). A positive change score indicates an increase, and a negative change score indicates a decrease in the level of the measured construct.

The use of such change scores is rational and widespread in both organisational and educational psychology. There are, however, influential methodologists who warn of

quantitative risks and recommend alternative options for detecting change (Thorndike & Thorndike-Christ, 2010; Willett, 1988).

Major parts of the methodological debate have focused on research environments that differ from the current thesis (Nesselroade, 1990; Nesselroade & Boker, 1994; Nesselroade & Ghisletta, 2003). An example is the distinction between estimating true development versus "increments/decrements only" (in this project, there was no intention to plot development per se). Types of scores targeted by criticism against change scores include derived scores such as percentiles (Russell, 2000a, 2000b). Again, the generation of change scores in this project avoided subtraction from percentiles. Further concern regarding levels of measurement, e.g. calculating change scores from nominal or ordinal input variables, is likewise irrelevant to the current project.

Objections regarding measurement error and reliability, however, warrant closer attention (Cronbach & Furby, 1970; Linn & Slinde, 1977; McNemar, 1958). Avoidance of simple change scores has been advocated as they may have greater measurement error and lower reliability than original scores obtained at times t_1 and t_2 (Cronbach & Furby, 1970; Lord, 1956). In concordance with True Score Theory of reliability, Willett (1988) warns about presuppositions of 'observed status' at t_1 and t_2 representing ‘true status’, before change scores are generated. Apart from the reliability levels of t_1 and t_2 scores, the reliability of the change score is a function of the correlation between the two ‘original’ scores. With higher correlation, reduction in reliability vis-à-vis the ‘original’ scores may be more pronounced (Feldt, 1967).

Even in situations where the t_1 and t_2 scores are equally highly reliable and have equal variances, the use of change scores has prompted caveats (Dimitrov & Rumrill Jr, 2003; Rogosa & Willett, 1983; Williams & Zimmerman, 1996a, 1996b). Criticisms levelled against using change scores revolve around test validity (Messick, 1981). Willett (1988) discusses a concern about 'equatability', namely, the consequences of re-measurements not having the same demonstrated level of construct validity. If a second measurement of the same purported variable fails to remain construct-valid, the change score will not have the same psychological meaning (Lord, 1956), and the assumption that measures at t_1 and t_2 have the same metric becomes non-defensible (Rogosa, Brandt, & Zimowski, 1982).

Authors such as Bereiter (1963) or Linn & Slinde (1977) allege that a change score cannot be simultaneously viewed as a valid and a reliable ‘measure’ of change. This so-called 'invalidity/ unreliability dilemma' has since been refuted (Rogosa et al., 1982; Rogosa & Willett, 1983; D. W. Zimmerman, Brotohusodo, & Williams, 1981; D. W. Zimmerman & Williams, 1982), with

several proofs that high reliability does not logically imply lower validity, and change scores cannot be branded necessarily unreliable.

Authors such as Williams and Zimmerman (1996a) posit that both the reliability criticism and the validity concern may partly rest on a misinterpretation of the notion of reliability as proposed in Classical Test Theory, and/or arbitrary assumptions about parameters in the reliability equation. Wittman (1988) argues that a 'condemnation' of utilising simple change scores may be premature. Willett (1989) risks an even bolder conclusion when stating, "the difference score has been demonstrated to be an intuitive, unbiased, and computationally simple measure of individual growth" (p.588).

A sound interpretation of these disputes is that the significance of change scores should be managed conservatively. One recommendation is to avoid viewing such differences as significant unless the lower score plus 1 SEM stays under the higher score minus 1 SEM (non-overlapping confidence bands; Cronbach & Furby, 1970). While reliability issues are worthy of attention, it has been suggested that in several crucial real-life settings, change scores can be sufficiently reliable for decision-making (Rogosa & Willett, 1983; D. W. Zimmerman & Williams, 1982). An example is group testing situations (Willett, 1988). More radical proposals involve alternatives to change scores, such as growth size estimates (Russell, 2000b), growth curves (Rogosa et al., 1982), covariance analysis (Willett & Sayer, 1994), configural frequency analysis (von Eye & Nesselroade, 1992), or procedures based on Item Response Theory (Prieler & Raven, 2009).

6.5 Considerations for future research

After examination of the findings and limitations of the current research, the following sections provide recommendations for future investigations.

6.5.1 Construct domain and choice of psychometric tools

The selection of self-concept and self-esteem from among the many self-referent constructs was a somewhat arbitrary decision. Future research could beneficially explore other constructs from the wider construct family, such as self-efficacy, self-image and self-knowledge.

6.5.2 Psychometrics

As with the selection of constructs above, the psychometric instruments chosen to measure these constructs represent only two choices from among many possible options. Future research could do more to critically compare the alternative standardised tests available for each self-perception construct. Empirical work could then be based on selection of other tests, or encompass a range of instruments simultaneously.

6.5.3 Proper sampling

The results from the current research should be replicated using a probabilistic sample of appropriate size (Field, 2009; Frankfort-Nachmias, 1996; Hansen et al., 1993). This sample can be selected from a tightly defined population, preferably one for which there is a sampling frame available (e.g. one defined via electoral rolls or similar databases). After conducting rigorous random sampling (possibly stratified), the replicated findings could then be generalised to the intended population. With regard to the population of organisations that utilise personality assessment for selection, sampling could also rely on a catalogue of organisations. The process could then be stratified by the type and size of organisation.

6.5.4 Restoring and expanding the original research design

Finally, key aspects of a longitudinal design involving several waves of measurement were conceived and described at the inception of the current project. While pursuing such a design requires a more comprehensive project at greater cost, the original design is worth restoring. After being able to reinstate the design and execute data collection accordingly, the hypotheses regarding change in self-perception over time could be operationalised in a more appropriate manner. Quantitative analyses involving path analysis, latent-variable models, or structural equation modeling could offer more robust opportunities for describing and providing evidence of change.

Appendices

Appendix 1: References

Abdi, H. (2007). The Bonferroni and Šidák corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103-107). Thousand Oaks, CA: Sage.

Acheson, S. K., & Thorpe, G. L. (2017). Test review of MMPI-2-RF. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The twentieth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Alexander, H. S., & McGarty, C. (2014). *Research methods and statistics in psychology* (2nd ed.). Los Angeles: Sage.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Allport, G. W. (1955). *Becoming: Basic considerations for a psychology of personality*. New Haven, CT: Yale University Press.

Ambrose, M. L., & Rosse, J. G. (2003). Procedural justice and personality testing: An examination of concern and typicality. *Group & Organization Management, 28*, 502-526. doi:10.1177/1059601103251234

Anderson, N. (2004). Editorial–The dark side of the moon: Applicant perspectives, negative psychological effects (NPEs), and candidate decision making in selection. *International Journal of Selection and Assessment, 12*(1 ‐ 2), 1-8. doi:10.1111/j.0965-075X.2004.00259.x

Arnold, B. R., & Smith, J. L. (2013). Methodologies for test translation and cultural equivalence. In F. A. Paniagua & A. M. Yamada (Eds.), *Handbook of multicultural mental health: assessment and treatment of diverse populations* (2nd ed., pp. 243-262). Amsterdam: Academic Press.

Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology, 69*(6), 1102-1112. doi:10.1037/0022-3514.69.6.1102

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*(2), 380-393. doi:10.1037/0022-3514.83.2.380

Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology, 27*, 519-533. doi:10.1111/j.1744-6570.1974.tb01173.x

Atlas, J. A., Gable, R. K., & Isonio, S. (1998). Test review of Self-Description Questionnaire--I, II, III. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Axford, S. N., & LoBello, S. G. (1998). Test review of Hogan Personality Inventory [Revised]. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Ayub, N. (2010). The relationship between self-concept and satisfaction with life among adolescents. *International Journal of Interdisciplinary Social Sciences, 5*(4), 81-92.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*(3), 533-548. doi:10.1037/a0016229

Bagley, C., Bolitho, F., & Bertrand, L. (2007). Norms and construct validity of the Rosenberg Self-Esteem Scale in Canadian high school populations: Implications for counselling. *Canadian Journal of Counselling and Psychotherapy/Revue canadienne de counseling et de psychothérapie, 31*(1), 82-92. doi:10.1037/t01038-000

Bajic, E. (2015, September 28). How the MBTI can help you build a stronger company. *Forbes*. Retrieved from https://www.forbes.com/sites/elenabajic/2015/09/28/how-the-mbti-can-help-you-build-a-stronger-company/ - 3a16f102d93c

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. doi:10.1037/0003-066X.37.2.122

Bandura, A. (1999). Social cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 154-196). New York: Guilford Press.

Barenbaum, N. B., & Winter, D. G. (2008). History of modern personality theory and research. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 3-26). New York: Guilford Press.

Barrett, P., Kline, P., Paltiel, L., & Eysenck, H. J. (1996). An evaluation of the psychometric properties of the concept 5.2 Occupational Personality Questionnaire. *Journal of Occupational and Organizational Psychology, 69*(1), 1-19. doi:10.1111/j.2044-8325.1996.tb00595.x

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

Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185 - 1203. doi:10.1037/0021-9010.90.6.1185

Bartram, D. (2007). Increasing validity with forced‐choice criterion measurement formats. *International Journal of Selection and Assessment, 15*(3), 263-272. doi:10.1111/j.1468-2389.2007.00386.x

Bartram, D., & Brown, A. (2005). *Five factor model (Big Five) OPQ32 report: OPQ32 technical manual supplement*. Thames Ditton (Surrey, United Kingdom): SHL Group.

Baumeister, R. F. (1997). Identity, self-concept, and self-esteem: The self lost and found. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 681-710). San Diego, CA: Academic Press.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 680-740). New York: McGraw-Hill.

Bell, B. S., Ryan, A. M., & Wiechmann, D. (2004). Justice expectations and applicant perceptions. *International Journal of Selection and Assessment, 12*(1‐2), 24-38. doi:10.1111/j.0965-075X.2004.00261.x

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1-62). New York: Academic Press.

Bembenutty, H. (2009). The last word: An interview with Herbert W. Marsh: The leading voice on self-concept, teaching effectiveness, and a force in quantitative analysis (Part 1). *Journal of Advanced Academics, 20*(3), 536-545. doi:10.1177/1932202X1102200207

Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota multiphasic personality inventory-2 restructured form: MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.

Benson, N., & Kluck, A. S. (2014). Test review of NEO Personality Inventory-3. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The nineteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, WI: University of Wisconsin Press.

Bergner, R. M. (2017). What is a person? What is the self? Formulations for a science of psychology. *Journal of Theoretical and Philosophical Psychology, 37*(2), 77-90. doi:10.1037/teo0000057

Black, J. (2000). Personality testing and police selection: Utility of the 'Big Five'. *New Zealand Journal of Psychology, 29*(1), 2-9.

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, L. S. Wrightsman, & F. M. Andrews (Eds.), *Measures of social psychological attitudes* (Vol. 1, pp. 115-160). San Diego, CA: Academic Press.

Boag, S. (2015). Personality assessment,'construct validity', and the significance of theory. *Personality and Individual Differences, 84*, 36-44. doi:10.1016/j.paid.2014.12.039

Bogart, L. (1967). No opinion, don't know, and maybe no answer. *Public Opinion Quarterly, 31*(3), 331-345. doi:10.1086/267533

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In (No editor's name printed) (Ed.), *Studi in onore del Professore Salvatore Ortu Carboni* (pp. 13-60). Rome, Italy: Without indication of publishing house. [Copy= Università degli Studi di Roma 'La Sapienza', Roma, Italy, # PUV-0266290; BNI 1935-10574.].

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8*, 3-62.

Boyce, C. J., Wood, A. M., & Powdthavee, N. (2013). Is personality fixed? Personality changes as much as 'variable' economic factors and more strongly predicts changes to life satisfaction. *Social Indicators Research, 111*(1), 287-305. doi:10.1007/s11205-012-0006-z

Boylan, R. V. (2014). *Decision criteria for selection of psychometric assessment instruments used in employee selection and development applications: A Delphi study [Abstract].* (D.Ed. Doctoral dissertation), North Carolina State University, Raleigh, NC. Ann Arbor, MI: Proquest.

British Psychological Society. (2009). *Occupational Personality Questionnaire (OPQ32): Summary of Test Review*. Retrieved from https://ptc.bps.org.uk/test-review/occupational-personality-questionnaire-opq32

Brown, A., & Bartram, D. (2009a). *Development and psychometric properties of OPQ32r: Supplement to the OPQ32 technical manual*. Thames Ditton (Surrey, United Kingdom): SHL Group.

Brown, A., & Bartram, D. (2009b). *The Occupational Personality Questionnaire revolution: Applying Item Response Theory to questionnaire design and scoring*. Retrieved from http://www.organisationalpsychology.nz/_content/The_Occupational_Personality_Questionnaire_Revolution.pdf

Brown, R., & Hattie, J. (1998). Test review of Tennessee Self-Concept Scale, Second Edition. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99-136). Washington, DC: American Psychological Association.

Buhs, E. S., & Scott, A. N. (2014). Test review of Self-Esteem Assessment. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The nineteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Burger, J. M. (2015). *Personality* (9th ed.). Stamford, CT: CENGAGE Learning.

Burke, E. (2008). Coaching with OPQ. In J. Passmore (Ed.), *Psychometrics in coaching: Using psychological and psychometric tools for development.* (pp. 87-114). London: Kogan Page.

Burr, J., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research. Vol. 1: Principles and structuring change* (pp. 3-34). New York: Academic Press.

Buss, D. M. (2008). Human nature and individual differences. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 29-60). New York: Guilford Press.

Butcher, J. N. (2010). Minnesota Multiphasic Personality Inventory. In I. B. Weiner & W. E. Craighead (Eds.), *Corsini Encyclopedia of Psychology* (pp. 1-3). Hoboken, NJ: Wiley.

Cairns, T. D., & Preziosi, R. C. (2014). Why HR leaders should encourage diversity of thought. *Employment Relations Today, 41*(2), 1-9. doi:10.1002/ert

Camara, W. J., & Merenda, P. F. (2000). Using personality tests in preemployment screening: Issues raised in Soroka v. Dayton Hudson Corporation. *Psychology, Public Policy, and Law, 6*(4), 1164-1186. doi:10.1037//1076-8971.6.4.1164

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. doi:10.1037/h0046016

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*(1), 141-156. doi:10.1037/0022-3514.70.1.141

Carducci, B. J. (2015). *The psychology of personality: Viewpoints, research, and applications* (3rd ed.). Hoboken, NJ: Wiley.

Carless, S. A. (2009). Psychological testing for selection purposes: a guide to evidence-based practice for human resource professionals. *The International Journal of Human Resource Management, 20*(12), 2517-2532. doi:10.1080/09585190903363821

Carrington-Rotto, P., & McLellan, M. J. (1995). Test review of Sixteen Personality Factor Questionnaire, Fifth Edition. In J. C. Conoley & J. C. Impara (Eds.), The twelfth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Cervone, D. (1991). The two disciplines of personality psychology. *Psychological Science, 2*(6), 371-377. doi:10.1111/j.1467-9280.1991.tb00169.x

Cervone, D., & Mischel, W. (2002). Personality science. In D. Cervone & W. Mischel (Eds.), *Advances in Personality Science* (pp. 1-26). New York: Guilford Press.

Cervone, D., & Pervin, L. A. (2013). *Personality: Theory and research* (12th ed.). Hoboken, NJ: Wiley.

Cervone, D., Shadel, W. G., & Jencius, S. (2001). Social-cognitive theory of personality assessment. *Personality and Social Psychology Review, 5*(1), 33-51. doi:10.1207/S15327957PSPR0501_3

Christiansen, N., & Tett, R. (2013). The long and winding road: An introduction to the Handbook of personality at work. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 1-7). New York: Brunner-Routledge.

Ciarrochi, J., & Bilich, L. (2006). *Acceptance and Commitment Therapy. Measures package*. Unpublished manuscript. School of Psychology. University of Wollongong. Wollongong (New South Wales, Australia).

Code of Ethics Review Group. (2012). *Code of ethics for psychologists working in Aotearoa/New Zealand*. Retrieved from http://www.psychologistsboard.org.nz/cms_show_download.php?id=237

Cohen, L., Leibu, O., Tanis, T., Ardalan, F., & Galynker, I. (2016). Disturbed self concept mediates the relationship between childhood maltreatment and adult personality

pathology. *Comprehensive Psychiatry, 68*, 186-192. doi:10.1016/j.comppsych.2016.04.020
Collins, L. M. (1996). Is reliability obsolete? A commentary on" Are simple gain scores obsolete?". *Applied Psychological Measurement, 20*(3), 289-292. doi:10.1177/014662169602000308
Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco, CA: W. H. Freeman and Company.
Cronbach, L. J., & Drenth, P. J. D. (Eds.). (1973). *Mental tests and cultural adaptation* (Vol. 7). The Hague, Netherlands: Mouton.
Cronbach, L. J., & Furby, L. (1970). How we should measure" change": Or should we? *Psychological Bulletin, 74*(1), 68-80. doi:10.1037/h0029382
Crumpton, N. L., & Farmer, R. F. (2010). Test review of Self-Perceptions of Adults. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), The eighteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.
Cuellar, I., & Paniagua, F. A. (2000). *Handbook of multicultural mental health: Assessment and treatment of diverse populations*. San Diego, CA: Academic Press.
Dachler, P. (1989). Selection and the organizational context. In P. Herriot & P. J. D. Drenth (Eds.), *Assessment and selection in organizations: Methods and practice for recruitment and appraisal* (pp. 45-69). Chichester, England: Wiley.
Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology*. Boston, MA: Allyn & Bacon.
Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35*(10), 867-881. doi:10.1037/0003-066X.35.10.867
De Cuyper, K., De Houwer, J., Vansteelandt, K., Perugini, M., Pieters, G., Claes, L., & Hermans, D. (2017). Using indirect measurement tasks to assess the self‐concept of personality: A systematic review and meta‐analyses. *European Journal of Personality, 31*(1), 8-41. doi:10.1002/per.2092
Dean, G. J., & Freeman, S. J. (2010). Test review of CPI 260. In J. C. Conoley & J. C. Impara (Eds.), The eighteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.
Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*(2), 79-86. doi:10.1007/s10111-003-0136-9
Deloitte Australia. (2011a). *Only skin deep? Re-examining the business case for diversity*. Retrieved from https://www.ced.org/pdf/Deloitte_-_Only_Skin_Deep.pdf
Deloitte Australia. (2011b). *Working in an ideological echo chamber? Diversity as a breakthrough strategy*. Retrieved from https://www.globalwomen.org.nz/assets/Uploads/resources/Leadership-resources-Deloitte-Working-in-an-Ideological-echo-chamber.pdf
Demo, D. H. (1992). The self-concept over time: Research issues and directions. *Annual Review of Sociology, 18*(1), 303-326. doi:10.1146/annurev.so.18.080192.001511
Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159-165.
Ellis-Hill, C. S., & Horn, S. (2000). Change in identity and self-concept: a new theoretical approach to recovery following a stroke. *Clinical Rehabilitation, 14*(3), 279-287. doi:10.1037/t03589-000
Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Employment Court Christchurch. (2013). Derek Wayne Gilbert v Transfield Services (New Zealand) Limited. Judgement of Chief Judge. NZEmpC 71 [29 April 2013]. Document # 46/10.

Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology, 41*(2), 232. doi:10.1037/0022-3514.41.2.232

Feldman, J. M. (1992). Constructive processes as a source of context effects in survey research: Explorations in self-generated validity. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 49-61). New York: Springer.

Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421-435. doi:10.1037/0021-9010.73.3.421

Feldt, L. S. (1967). A note on the use of confidence bands to evaluate the reliability of a difference between two scores. *American Educational Research Journal, 4*(2), 139-145.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of consulting and clinical psychology, 43*(4), 522-527. doi:10.1037/h0076760

Ferguson, E., & Lievens, F. (2017). Future directions in personality, occupational and medical selection: Myths, misunderstandings, measurement, and suggestions. *Advances in Health Sciences Education, 22*(2), 387-389. doi:10.1007/s10459-016-9751-0

Ferguson, E., Payne, T., & Anderson, N. (1994). Occupational personality assessment: Theory, structure and psychometrics of the OPQ FMX5-Student. *Personality and Individual Differences, 17*(2), 217-225. doi:10.1016/0191-8869(94)90028-0

Festinger, L. (1957). *A theory of cognitive dissonance*. London: Tavistock.

Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: Sage.

Forbey, J. D., & Ben-Porath, Y. S. (2007). A comparison of the MMPI-2 restructured clinical (RC) and clinical scales in a substance abuse treatment sample. *Psychological Services, 4*(1), 46-58. doi:10.1037/1541-1559.4.1.46

Foster, G. C., Min, H., & Zickar, M. J. (2017). Review of Item Response Theory practices in organizational research: Lessons learned and paths forward. *Organizational Research Methods, 20*(3), 465-486. doi:10.1177/1094428116689708

Frankfort-Nachmias, C. (1996). *Research methods in the social sciences* (5th ed.). London: Edward Arnold.

Funder, D. C. (2010). *The personality puzzle* (5th ed.). New York: W.W. Norton.

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2016). *Human resource selection* (8th ed.). Boston, MA: CENGAGE Learning.

Gecas, V., Calonico, J. M., & Thomas, D. L. (1974). The development of self-concept in the child: Mirror theory versus model theory. *The Journal of Social Psychology, 92*(1), 67-76. doi:10.1080/00224545.1974.9923073

Gergen, K. J. (2011). The self as social construction. *Psychological Studies, 56*(1), 108-116. doi:10.1007/s12646-011-0066-1

Gerras, S. J., & Wong, L. (2016). Moving beyond the MBTI: The Big Five and leader development. *Military Review, 96*(2), 54-57.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694-734. doi:10.5465/AMR.1993.9402210155

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84-96. doi:10.1016/j.jrp.2005.08.007

Goldsmith, R. E. (1986). Dimensionality of the Rosenberg self-esteem scale. *Journal of social Behavior and Personality, 1*(2), 253-264.
Gordon, E., & Webb, R. (2014). A practical approach to psychometric profiling; delivering business value. *Human Resources, 18,* 14-15.
Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 23*(5), 443-451. doi:10.1177/0146167297235001
Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences, 35*(6), 1241-1254. doi:10.1016/S0191-8869(02)00331-8
Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603-618. doi:10.1037/0003-066X.35.7.603
Greuter, M. A., & Algera, J. A. (1989). Criterion development and job analysis. In P. Herriot & P. J. D. Drenth (Eds.), *Assessment and selection in organizations: Methods and practice for recruitment and appraisal* (pp. 143-181). Chichester, England: Wiley.
Grice, J. W. (2015). From means and variances to persons and patterns. *Frontiers In Psychology, 6*, 1-12. doi:10.3389/fpsyg.2015.01007
Griffiths, R. A., Beumont, P. J., Giannakopoulos, E., Russell, J., Schotte, D., Thornton, C., . . . Varano, P. (1999). Measuring self‐esteem in dieting disordered patients: The validity of the Rosenberg and Coopersmith contrasted. *International Journal of Eating Disorders, 25*(2), 227-231.
Grubbs, J. B., Wilt, J., Stauner, N., Exline, J. J., & Pargament, K. I. (2016). Self, struggle, and soul: Linking personality, self-concept, and religious/spiritual struggle. *Personality and Individual Differences, 101*, 144-152. doi:10.1016/j.paid.2016.05.365
Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions*. New York: Erlbaum.
Hagborg, W. J. (1993). The Rosenberg Self‐Esteem scale and Harter's Self‐Perception profile for adolescents: a concurrent validity study. *Psychology in the Schools, 30*(2), 132-136. doi:10.1002/1520-6807(199304)30:2<132::AID-PITS2310300205>3.0.CO;2-Z
Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1993). *Sample survey methods and theory* (Wiley classics library ed.). New York: Wiley.
Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley, CA: University of California Press.
Harris, N. J., Toulson, P. K., & Livingston, E. M. (1996). New Zealand personnel consultants and the selection process. *Asia Pacific Journal of Human Resources, 34*(2), 71-87. doi:10.1177/103841119603400208
Harter, S. (1996). Historical roots of contemporary issues. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 1-37). New York: Wiley.
Hartley, J. (2014). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology, 14*(1), 83-86. doi:10.1016/S1697-2600(14)70040-7
Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: Erlbaum.
Hattie, J., & Marsh, H. W. (1996). Future directions in self-concept research. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 421-462). New York: Wiley.
Hau, K.-T., Kong, C.-K., & Marsh, H. W. (2003). Chinese Self-Description Questionnaire. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self research* (pp. 49-66). Greenwich, CT: Information Age Publishing.
Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*(6), 895-910. doi:10.1037/0022-3514.60.6.895

Hendy, N. T. (2017). Forced-choice personality measures and academic dishonesty: A comparative study. *Journal of Academic Ethics*, 1-14. doi:10.1007/s10805-017-9280-3

Herbert, H. S., Manjula, M., & Philip, M. (2013). Resilience and factors contributing to resilience among the offsprings of parents with schizophrenia. *Psychological Studies, 58*(1), 80-88. doi:10.1007/s12646-012-0168-4

Hess, A. K., & Lanning, K. (2003). Test review of Myers-Briggs Type Indicator(r) Step II (Form Q). In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), The fifteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Highhouse, S., Doverspike, D., & Guion, R. M. (2016). *Essentials of personnel assessment and selection* (2nd ed.). New York: Routledge.

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

Hogan, R. (1982). A socioanalytic theory of personality. In R. A. Dienstbier, M. M. Page, L. Pervin, R. Hogan, S. Epstein, N. S. Endler, D. J. Bem, W. Mischel, & P. K. Peake (Eds.), *Nebraska symposium on motivation: Personality - current theory and research* (pp. 55-89). Lincoln, NE: University of Nebraska Press.

Hogan, R., & Blickle, G. (2013). Socioanalytic theory. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 53-70). New York: Brunner-Routledge.

Hogan, R., & Foster, J. (2016). Rethinking personality. *International Journal of Personality Psychology, 2*(1), 37-43.

Hogan, R., Hogan, J., & Warrenfeltz, R. (2007). *The Hogan guide: Interpretation and uses of Hogan inventories*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Roberts, B. W. (2001). Introduction: Personality and industrial and organizational psychology. In B. Roberts & R. Hogan (Eds.), Personality psychology in the workplace (1st ed., pp. 3-16). Washington, DC: American Psychological Association.

Hough, L. M. (1998). Personality at work: Issues and evidence. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-166). Mahwah, NJ: Erlbaum.

Hough, L. M., & Connelly, B. S. (2013). Personality measurement and use in industrial and organizational psychology. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology* (Vol. 1, pp. 501-531). Washington, DC: American Psychological Association.

Hough, L. M., & Dilchert, S. (2010). Personality: Its measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 299-319). New York: Routledge.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future--Remembering the past. *Annual Review of Psychology, 51*(1), 631-664. doi:10.1146/annurev.psych.51.1.631

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology, 1*(3), 272-290. doi:10.1111/j.1754-9434.2008.00057.x

Hudson, N. W., & Fraley, R. C. (2015). Volitional personality trait change: Can people choose to change their personality traits? *Journal of Personality and Social Psychology, Advance online publication*, 1-18. doi:10.1037/pspp0000021

Humphreys, L. G. (1996). Linear dependence of gain scores on their components imposes constraints on use and interpretation: Comment on" Are simple gain scores obsolete?". *Applied Psychological Measurement, 20*(3), 293-294.

Hutchinson, T. P. (1991). *Ability, partial information, guessing: Statistical modelling applied to multiple-choice tests*. Adelaide, Australia: Rumsby Scientific Publishing.

Imus, A., & Ryan, A. (2005). Relevance and rigor in research on the applicant's perspective: In pursuit of pragmatic science. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 291-305). Oxford, England: Blackwell.

Ion, A., & Iliescu, D. (2017). The measurement equivalence of personality measures across high-and low-stake test taking settings. *Personality and Individual Differences, 110*, 1-6. doi:10.1016/j.paid.2017.01.008

Jeanneret, P. R., & Zedeck, S. (2010). Professional guidelines/standards. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 593-625). New York: Routledge.

Johnson, E. K. (2000). The practice of human resource management in New Zealand: strategic and best practice? *Asia Pacific Journal of Human Resources, 38*(2), 69-83.

Johnson, J. A. (1981). The" self-disclosure" and" self-presentation" views of item response dynamics and personality scale validity. *Journal of Personality and Social Psychology, 40*(4), 761-769. doi:10.1037/0022-3514.40.4.761

Johnson, J. W., & Oswald, F. L. (2010). Test administration and the use of test scores. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 151-169). New York: Routledge.

Jones, E. E., Rhodewalt, F., Berglas, S., & Skelton, J. A. (1981). Effects of strategic self-presentation on subsequent self-esteem. *Journal of Personality and Social Psychology, 41*(3), 407. doi:10.1037/0022-3514.41.3.407

Jones, J. W. (1991). Assessing privacy invasiveness of psychological test items: Job relevant versus clinical measures of integrity. *Journal of Business and Psychology, 5*(4), 531-535.

Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K., & Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and Likert scale versions of a personality instrument. *International Journal of Selection and Assessment, 23*(1), 92-97. doi:10.1111/ijsa.12098

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*(1), 80-92. doi:10.1037//0021-9010.86.1.80

Judge, T. A., Erez, A., & Bono, J. E. (1998). The power of being positive: The relation between positive self-concept and job performance. *Human performance, 11*(2-3), 167-187. doi:10.1080/08959285.1998.9668030

Jung, C. G. (1923). *Psychological types*. London: Routledge.

Kam, C. D., Wilking, J. R., & Zechmeister, E. J. (2007). Beyond the "narrow data base": Another convenience sample for experimental research. *Political Behavior, 29*(4), 415-440. doi:10.1007/s11109-007-9037-6

Kantrowitz, T. (2014). *2014 Global assessment trends report*. Retrieved from Alpharetta, GA:

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues* (9th ed.). Belmont, CA: CENGAGE Learning.

Kehoe, J. F., & Murphy, K. (2010). Current concepts of validity, validation and generalizability. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 99-123). New York: Routledge.

Keith, L. K., & Bracken, B. A. (1996). Self-concept instrumentation: A historical and evaluative review. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 91-170). New York: Wiley.

Kihlstrom, J. F., & Klein, S. B. (1997). Self-knowledge and self-awareness. *Annals Of The New York Academy Of Sciences, 818*, 4-17.

Klehe, U. C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *International Journal of Selection and Assessment, 12*(4), 327-342. doi:10.1111/j.0965-075X.2004.00288.x

Klimstra, T. A., Bleidorn, W., Asendorpf, J. B., van Aken, M. A. G., & Denissen, J. J. A. (2013). Correlated change of Big Five personality traits across the lifespan: A search for

determinants. *Journal of Research in Personality, 47*(6), 768-777. doi:10.1016/j.jrp.2013.08.004
Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, DC: American Psychological Association.
Kohn, M. L., & Schooler, C. (1969). Class, occupation, and orientation. *American Sociological Review, 34*(5), 659-678. doi:10.2307/2092303
Krause, M. S. (2013). The data analytic implications of human psychology's dimensions being ordinally scaled. *Review of General Psychology, 17*(3), 318-325. doi:10.1037/a0032292
Kravitz, D. A., Stinson, V., & Chavez, T. L. (1996). Evaluations of tests used for making selection and promotion decisions. *International Journal of Selection and Assessment, 4*(1), 24-34. doi:10.1111/j.1468-2389.1996.tb00045.x
Kristof, A. L. (1996). Person‐organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49*(1), 1-49.
Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*(2), 281-342. doi:10.1111/j.1744-6570.2005.00672.x
Lamiell, J. T. (2000). A periodic table of personality elements? The" Big Five" and trait" psychology" in critical perspective. *Journal of Theoretical and Philosophical Psychology, 20*(1), 1-24. doi:10.1037/h0091211
Lamiell, J. T. (2007). On sustaining critical discourse with mainstream personality investigators: Problems and prospects. *Theory & Psychology, 17*(2), 169-185. doi:10.1177/0959354307075041
Lamiell, J. T. (2010). Why was there no place for personalistic thinking in 20th century psychology? *New Ideas in Psychology, 28*(2), 135-142. doi:doi:10.1016/j.newideapsych.2009.02.002
Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology, 31*(1), 65-71. doi:10.1016/j.newideapsych.2011.02.0 09
Larsen, R. J., Buss, D. M., & Wismeijer, A. (2013). *Personality psychology: Domains of knowledge about human nature* (5th ed.). New York: McGraw-Hill.
Lauver, K. J., & Kristof-Brown, A. L. (2001). Distinguishing between employees' perceptions of person–job and person–organization fit. *Journal of vocational behavior, 59*(3), 454-470. doi:10.1006/jvbe.2001.1807
Lawler, E. E., III. (2007). Why HR practices are not evidence-based. *Academy of Management Journal, 50*(5), 1033-1036. doi:10.5465/AMJ.2007.27155013
Leung, S. A., & Porchea, S. F. (2014). Test review of OPQ32. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The nineteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.
Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47*(1), 121-150.
Lord, F. M. (1956). The measurement of growth. *ETS Research Report Series*(1), 1-26.
Lu, H. J., & Chang, L. (2011). The association between self-deception and moral self-concept as functions of self-consciousness. *Personality and Individual Differences, 51*(7), 845-849. doi:10.1016/j.paid.2011.07.014
MacInnes, D. L. (2006). Self‐esteem and self‐acceptance: An examination into their relationship and their effect on psychological health. *Journal of Psychiatric and Mental Health Nursing, 13*(5), 483-489. doi:10.1111/j.1365-2850.2006.00959.x
Maehr, M. (2005). Preface. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self research: New frontiers for self research* (Vol. 2, pp. vii-viii). Greenwich, CT: Information Age Publishing.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*(2), 63-78. doi:10.1037/0022-3514.35.2.63
Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology, 51*(4), 858.
Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*(1), 299-337.
Marsella, A. J., & Leong, F. T. (1995). Cross-cultural issues in personality and career assessment. *Journal of Career Assessment, 3*(2), 202-218.
Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational psychology review, 2*(2), 77-172. doi:10.1007/BF01322177
Marsh, H. W., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133-163. doi:10.1111/j.1745-6916.2006.00010.x
Marsh, H. W., Craven, R., & McInerney, D. M. (2003). International advances in self research: Speaking to the future. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self research* (pp. 3-16). Greenwich, CT: Information Age Publishing.
Marsh, H. W., Craven, R., & McInerney, D. M. (2005). Overview: New frontiers for self research. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self research: New frontiers for self research* (Vol. 2, pp. 3-14). Greenwich, CT: Information Age Publishing.
Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 38-90). New York: Wiley.
Marsh, H. W., & O'Neill, R. (1984). Self description questionnaire III: the construct validity of multidimensional self‐concept ratings by late adolescents. *Journal of Educational Measurement, 21*(2), 153-174. doi:10.1111/j.1745-3984.1984.tb00227.x
Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational psychologist, 20*(3), 107-123. doi:10.1207/s15326985ep2003_1
Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality, 74*(2), 403-456. doi:10.1111/j.1467-6494.2005.00380.x
Martin, L. S., Oades, L. G., & Caputi, P. (2013). What is personality change coaching and why is it important? A response to Martin, Oades & Caputi. *International Coaching Psychology Review, 8*(1), 101-103.
Matthews, G., & Stanton, N. (1994). Item and scale factor analyses of the Occupational Personality Questionnaire. *Personality and Individual Differences, 16*(5), 733-743. doi:10.1016/0191-8869(94)90214-3
Matthews, G., Stanton, N., Graham, N. C., & Brimelow, C. (1990). A factor analysis of the scales of the Occupational Personality Questionnaire. *Personality and Individual Differences, 11*(6), 591-596. doi:10.1016/0191-8869(90)90042-P
Maul, A. (2013). On the ontology of psychological attributes. *Theory & Psychology, 0*(0), 1-18. doi:10.1177/0959354313506273
McAdams, D. P. (2009). *The person: An introduction to the science of personality psychology* (5th ed.). Hoboken, NJ: Wiley.
McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*(3), 204-217. doi:10.1037/0003-066X.61.3.204

McCrae, R. R., & Costa, P. T. (2008). The five factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159-181). New York: Guilford Press.
McFarland, L. (2013). Applicant reactions to personality tests. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 281-298). New York: Brunner-Routledge.
McNemar, Q. (1958). On growth measurement. *Educational and Psychological Measurement, 18*(1), 47-55.
Messick, S. (1981). Denoting the base-free measure of change. *Psychometrika, 46*, 215-217. doi:10.1007/BF02293901
Messick, S. (1990). Validity of test interpretation and use. *ETS Research Report Series, 1990*(1), 1487-1495. doi:10.1002/j.2333-8504.1990.tb01343.x
Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355-383. doi:10.1111/j.2044-8295.1997.tb02641.x
Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology, 10*(5), 639-667.
Michell, J. (2004). Item response models, pathological science and the shape of error: Reply to Borsboom and Mellenbergh. *Theory & Psychology, 14*(1), 121-129. doi:10.1177/0959354304040201
Michell, J. (2008). Is Psychometrics Pathological Science? *Measurement, 6*(1/2), 7-24. doi:10.1080/15366360802035489
Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. *Frontiers In Psychology, 3*, 1-34. doi:10.3389/fpsyg.2012.00261
Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology, 31*(1), 13-21. doi:doi:10.1016/j.newideapsych.2011.02.0 04
Ministry of Health. (2004). *Ethnicity data protocols for the health and disability sector*. Retrieved from Wellington, New Zealand:
Mischel, W., Shoda, Y., & Ayduk, O. (2008). *Introduction to personality: Toward an integrative science of the person* (8th ed.). New York: Wiley.
Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112-117. doi:10.1111/j.1467-8721.2009.01619.x
Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729. doi:10.1111/j.1744-6570.2007.00089.x
Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research, 20*(1), 46-61. doi:10.2307/2489199
Murphy, K., Deckert, P. J., & Hunter, S. T. (2013). What personality does and does not predict and why. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 633-650). New York: Brunner-Routledge.
Nesselroade, J. R. (1990). Adult personality development: Issues in assessing constancy and change. In A. I. Rabin, R. A. Zucker, R. A. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 41-85). New York: Springer.
Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. Heatherton & J. Weinberger (Eds.), *Can personality change?* (pp. 121-147). Washington, DC: American Psychological Association.
Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the life span. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 317-337). Boston: Kluwer Academic Publishers.
Nestor, P., & Schutt, R. K. (2015). *Research methods in psychology: Investigating human behavior* (2nd ed.). Thousand Oaks, CA: Sage.
Neukrug, E. (2015). *Essentials of testing and assessment: A practical guide to counselors, social workers, and psychologists* (3rd ed.). Stamford, CT: CENGAGE Learning.

Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management, 24*(1), 28-45. doi:10.1177/1059601199241003

New Zealand Council of Education Research. (2017). NZCER Psychological Test Services Registration. Retrieved from http://www.nzcer.org.nz/system/files/PTS Registration Levels.pdf

New Zealand Psychologists Board. (2015). *Guidelines on the use of psychometrics tests*. Retrieved from http://www.psychologistsboard.org.nz/cms_show_download.php?id=354

Newmark, J. (1997). *Statistics and probability in modern life* (6th ed.). Fort Worth, TX: Saunders College.

Nicholson, I. A. (2003). *Inventing personality: Gordon Allport and the science of selfhood*. Washington, DC: American Psychological Association.

O'Neill, T. A., Lewis, R. J., Law, S. J., Larson, N., Hancock, S., Radan, J., . . . Carswell, J. J. (2017). Forced-choice pre-employment personality assessment: Construct validity and resistance to faking. *Personality and Individual Differences, 115*, 120-127. doi:10.1016/j.paid.2016.03.075

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

Ones, D. S., & Viswesvaran, C. (2001). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In B. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (1st ed., pp. 31-39). Washington, DC: American Psychological Association.

Pajares, F., & Schunk, D. H. (2005). Self-efficacy and self-concept beliefs: Jointly contributing to the quality of human life. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self research: New frontiers for self research* (Vol. 2, pp. 95-122). Greenwich, CT: Information Age Publishing.

Paunonen, S. V., & Hong, R. Y. (2015). On the properties of personality traits. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology* (Vol. 4, pp. 233-259). Washington, DC: American Psychological Association.

Pearlman, K., & Sanchez, J. I. (2010). Work analysis. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 73-98). New York: Routledge.

Penney, L. M., David, E., & Witt, L. A. (2011). A review of personality and performance: Identifying boundaries, contingencies, and future research directions. *Human Resource Management Review, 21*(4), 297-310. doi:10.1016/j.hrmr.2010.10.005

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal, 316*, 1236-1238.

Perugini, M., & Banse, R. (2007). Personality, implicit self-concept and automaticity. *European Journal of Personality, 21*(3), 257-261. doi:10.1002/per.637

Peterson, R. A. (2005). Response construction in consumer behavior research. *Journal of Business Research, 58*(3), 348-353. doi:10.1016/S0148-2963(03)00100-0

Phylactou, C. A. (2000). *Correlates of self esteem in preadolescence.* (Doctor of Philosophy Doctoral dissertation), University of Surrey. Core database.

Ployhart, R. E., & Harold, C. M. (2004). The Applicant Attribution‐Reaction Theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment, 12*(1‐2), 84-98. doi:10.1111/j.0965-075X.2004.00266.x

Ployhart, R. E., & Ryan, A. M. (1997). Toward an explanation of applicant reactions: An examination of organizational justice and attribution frameworks. *Organizational Behavior and Human Decision Processes, 72*(3), 308-335. doi:10.1006/obhd.1997.2742

Prewett, M. S., Tett, R., & Christiansen, N. (2013). A review and comparison of 12 personality inventories on key psychometric characteristics. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 191-225). New York: Brunner-Routledge.

Prieler, J. A., & Raven, J. J. (2009). The measurement of change in groups and individuals with particular reference to the value of change scores: A new IRT-based methodology for the assessment of treatment effects. *International Journal of Educational and Psychological Assessment, 3*, 38-74.

Prien, E. P., Schippmann, J. S., & Prien, K. O. (2003). *Individual assessment: As practiced in industry and consulting*. Mahwah, NJ: Erlbaum.

Psychological Testing Centre, B. (2016). Occupational Personality Questionnaire (OPQ32): Summary of test review. Retrieved from https://ptc.bps.org.uk/test-review/occupational-personality-questionnaire-opq32

Psytech International. (2016). 15FQ+: Fifteen Factor Questionnaire technical manual. Retrieved from http://www.psytech.com/Content/TechnicalManuals/EN/15FQplusman.pdf

Public Service Association. (2013a). Government departments spend $1.5 million on psychometric testing [Press release]. Retrieved from https://www.psa.org.nz/media/releases/government-departments-spend-1-5-million-on-psychometric-testing/

Public Service Association. (2013b, September 15). Testing personalities. *PSA Journal*. Retrieved from https://www.psa.org.nz/media/news/testing-personalities/

Public Service Association. (2017). PSA taking Inland Revenue to court over unnecessary psychometric testing [Press release]. Retrieved from https://www.psa.org.nz/media/releases/psa-taking-inland-revenue-to-court-over-unnecessary-psychometric-testing/

Pullmann, H., & Allik, J. (2000). The Rosenberg Self-Esteem Scale: Its dimensionality, stability and personality correlates in Estonian. *Personality and Individual Differences, 28*(4), 701-715. doi:10.1016/S0191-8869(99)00132-4

Quenk, N. L. (2009). *Essentials of Myers-Briggs Type Indicator assessment* (2nd ed.). Hoboken, NJ: Wiley.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180.

Reynierse, J. H. (2013). Type theory revisited: Implications of complementary opposites for the five-factor model of personality and organizational interventions. In N. Christiansen & R. Tett (Eds.), *Handbook of personality at work* (pp. 356-391). New York: Brunner-Routledge.

Rhodewalt, F., & Agustsdottir, S. (1986). Effects of self-presentation on the phenomenal self. *Journal of Personality and Social Psychology, 50*(1), 47. doi:10.1037/0022-3514.50.1.47

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3-25. doi:10.1037/0033-2909.126.1.3

Robertson, I., & Kinder, A. (1993). Personality and job competences: The criterion‐related validity of some personality variables. *Journal of Occupational and Organizational Psychology, 66*(3), 225-244. doi:10.1111/j.2044-8325.1993.tb00534.x

Robertson, I., & Smith, M. (1989). Personnel selection methods. In M. Smith & I. Robertson (Eds.), *Advances in Selection and Assessment* (pp. 89-112). Great Britain: Wiley.

Robins, R. W., Tracy, J. L., & Trzesniewski, K. H. (2008). Naturalizing the self. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 421-447). New York: Guilford Press.

Roe, R. A. (2005). The design of selection systems: Context, principles, issues. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 73-97). Oxford, England: Blackwell.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748. doi:10.1037/0033-2909.92.3.726

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability the difference score in the measurement of change. *Journal of Educational Measurement, 20*(4), 335-343. doi:10.1111/j.1745-3984.1983.tb00211.x

Rose, N. S. (1996). *Inventing our selves: Psychology, power, and personhood*. Cambridge, England: Cambridge University Press.

Rosenberg, M. (1965). Rosenberg Self Esteem Scale [Database record]. Retrieved from PsycTESTS

Rosenberg, M. (1971). *Black and white self-esteem: The urban school child*. Washington, DC: American Sociological Association.

Rosenberg, M. (1986). *Conceiving the self* (Reprint ed.). New York: Basic Books.

Rosenberg, M. (1989). *Society and the adolescent self-image* (Rev. ed.). Middletown, CT: Wesleyan University Press.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60*(1), 141-156. doi:10.2307/2096350

Rosse, J. G., Miller, J. L., & Stecher, M. D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology, 79*(6), 987-992. doi:10.1037/0021-9010.79.6.987

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*(2), 155-180. doi:10.1016/j.hrmr.2006.03.004

Russell, M. (2000a). Summarizing change in test scores: Shortcomings of three common methods. *Practical Assessment, Research & Evaluation, 7*(5). Retrieved from PAREonline.net website:

Russell, M. (2000b). Using expected growth size estimates to summarize test score changes. *Practical Assessment, Research & Evaluation, 7*(6). Retrieved from PAREonline.net website:

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*(2), 359-391.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*(3), 565-606. doi:10.1177/014920630002600308

Ryden, M. B. (1978). An adult version of the Coopersmith Self-Esteem Inventory: Test-retest reliability and social desirability. *Psychological Reports, 43*(3), 1189-1190. doi:10.2466/pr0.1978.43.3f.1189

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*(2), 149-174. doi:10.1002/hrm.10029

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science, 9*(5), 538-551. doi:10.1177/1745691614543972

Salgado, J. F. (1996). Personality and job competencies: A comment on the Robertson & Kinder (1993) study. *Journal of Occupational and Organizational Psychology, 69*(4), 373-375.

Salgado, J. F., & De Fruyt, F. (2005). Personality in personnel selection. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 174-198). Oxford, England: Blackwell.

Samuda, R. J. (1998). *Psychological testing of American minorities: Issues and consequences* (2nd ed.). Thousand Oaks, CA: Sage.

Sanders, K., van Riemsdijk, M., & Groen, B. (2008). The gap between research and practice: A replication study on the HR professionals' beliefs about effective human resource practices. *The International Journal of Human Resource Management, 19*(10), 1976-1988. doi:10.1080/09585190802324304

Saville, P., Holdsworth, R., Nyfield, G., Cramp, L., & Mabey, W. (1984). *The Occupational Personality Questionnaires (OPQ)*. London: Saville & Holdsworth (UK) Ltd.

Saville, P., Sik, G., Nyfield, G., Hackston, J., & Maclver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in the measurement of job competencies across time and in separate organisations. *Applied Psychology, 45*(3), 243-262. doi:10.1111/j.1464-0597.1996.tb00767.x

Saville, P., & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology, 64*(3), 219-238. doi:10.1111/j.2044-8325.1991.tb00556.x

Schlenker, B. R., Dlugolecki, D. W., & Doherty, K. (1994). The impact of self-presentations on self-appraisals and behavior: The power of public commitment. *Personality and Social Psychology Bulletin, 20*(1), 20-33. doi:10.1177/0146167294201002

Schmidt, F. L., Oh, I.-S., & Shaffer, J. A. (2016). *The validity and utliity of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings*. Working paper.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*(4), 623-642. doi:10.1037/0022-3514.89.4.623

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

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 91*(3), 613-621. doi:10.1037/0021-9010.91.3.613

Sewell, T. E., Peterson, C., & Austin, J. T. (1985). Test review of Coopersmith Self-Esteem Inventories. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/.

Shadel, W. G. (2004). Introduction to the special series what can personality science offer cognitive-behavioral therapy and research? *Behavior Therapy, 35*(1), 101-111. doi:10.1016/S0005-7894(04)80006-6

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta‐analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*(3), 445-494. doi:10.1111/j.1744-6570.2012.01250.x

Shapurian, R., Hojat, M., & Nayerahmadi, H. (1987). Psychometric characteristics and dimensionality of a Persian version of Rosenberg Self-esteem Scale. *Perceptual and Motor Skills, 65*(1), 27-34.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407-441. doi:10.3102/00346543046003407

Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology, 73*(6), 1380-1393. doi:10.1037/0022-3514.73.6.1380

SHL Limited (formerly Saville & Holdsworth). (2013). *OPQ Profile. Technical report*. Thames Ditton (Surrey, United Kingdom): CEB/SHL Talent Measurement. Retrieved from https://www.shl.hu/sites/.

Silvia, P. J., & Duval, T. S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and Social Psychology Review, 5*(3), 230-241.
Simmons, C. J., Bickart, B. A., & Lynch Jr, J. G. (1993). Capturing and creating public opinion in survey research. *Journal of Consumer Research, 20*(2), 316-329.
Slaney, K. L. (2017). *Validating psychological constructs: Historical, philosophical, and practical dimensions*. London: Palgrave Macmillan UK.
Slaney, K. L., & Garcia, D. A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology, 35*(4), 244-259. doi:10.1037/teo0000025
Slaney, K. L., & Racine, T. P. (2013). What's in a name? Psychology's ever evasive construct. *New Ideas in Psychology, 31*(1), 4-12. doi:10.1016/j.newideapsych.2011.02.003
Smalley, R., & Stake, J. E. (1992). *The gender role/self-concept link: Looking beyond the college sophomore*. Paper presented at the 100th Annual Convention of the American Psychological Association, Washington, DC.
Smith, M., & George, D. (1994). Selection methods. In C. L. Cooper & I. T. Robertson (Eds.), *Key reviews in managerial psychology : concepts and research for practice* (pp. 54-96). Chichester, England: Wiley.
Specht, J., Bleidorn, W., Denissen, J. J., Hennecke, M., Hutteman, R., Kandler, C., . . . Zimmermann, J. (2014). What drives adult personality development? A comparison of theoretical perspectives and empirical evidence. *European Journal of Personality, 28*(3), 216-230. doi:10.1002/per.1966
Stake, J. E. (1985). Predicting reactions to everyday events from measures of self-esteem. *Journal of Personality, 53*(4), 530-542. doi:10.1111/1467-6494.ep8970358
Stake, J. E. (1992). Gender differences and similarities in self-concept within everyday life contexts. *Psychology of Women Quarterly, 16*(3), 349-363. doi:10.1111/j.1471-6402.1992.tb00259.x
Stake, J. E. (1994). Development and validation of the Six-Factor Self-Concept Scale for adults. *Educational and Psychological Measurement, 54*(1), 56-72. doi:10.1177/0013164494054001006
Stake, J. E., & Eisele, H. (2010). Gender and personality. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology* (pp. 19-40). New York: Springer.
Stake, J. E., Huff, L., & Zand, D. (1995). Trait self-esteem, positive and negative events, and event-specific shifts in self-evaluation and affect. *Journal of Research in Personality, 29*, 223-241. doi:10.1006/jrpe.1995.1013
Stanton, N. (1994). Item and scale factor analyses of the Occupational Personality Questionnaire. *Personality and Individual Differences, 16*(5), 733-743. doi:10.1016/0191-8869(94)90214-3
Statistics New Zealand. (2017). *Labour market statistics: June 2017 quarter*. Retrieved from: http://archive.stats.govt.nz/browse_for_stats/income-and-work/employment_and_unemployment/LabourMarketStatistics_HOTPJun17qtr/Commentary.aspx
Suzuki, L., Meller, P. J., & Ponterotto, J. G. (1996). Multicultural assessment: Current trends and future directions. In L. Suzuki, P. J. Meller, & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological and educational applications* (pp. 666-671). San Francisco, CA: Jossey-Bass.
Swann, W. B., & Ely, R. J. (1984). A battle of wills: self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology, 46*(6), 1287-1302. doi:10.1037//0022-3514.46.6.1287
Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices New Zealand organisations and recruitment firms. *New Zealand Journal of Psychology, 31*(1), 8-18.

Taylor, P., Mills, A., & O'Driscoll, M. (1993). Personnel selection methods used by New Zealand organisations and personnel consulting firms. *New Zealand Journal of Psychology, 22*(1), 19-31.

Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*(4), 967-993. doi:10.1111/j.1744-6570.2007.00098.x

The Economist. (2013, April 6). Emotional breakdown: Can leaders be identified by psychometrics? *The Economist*. Retrieved from http://www.economist.com/news/business/21575817-can-leaders-be-identified-psychometrics-emotional-breakdown

The Myers and Briggs Foundation. (2016). MBTI Basics. Retrieved from http://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/home.htm?bhcp=1

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Prentice Hall.

Tice, D. M. (1992). Self-concept change and self-presentation: the looking glass self is also a magnifying glass. *Journal of Personality and Social Psychology, 63*(3), 435-451.

Valentine, E. R. (1992). *Conceptual issues in psychology* (2nd ed.). London: Routledge.

Vallieres, E. F., & Vallerand, R. J. (1990). Traduction et validation canadienne ‐ française de l'échelle de l'estime de soi de Rosenberg. *International journal of psychology, 25*(2), 305-316.

Van Vianen, A. E., Taris, R., Scholten, E., & Schinkel, S. (2004). Perceived fairness in personnel selection: Determinants and outcomes in different stages of the assessment procedure. *International Journal of Selection and Assessment, 12*(1/2), 149-159. doi:10.1111/j.0965-075X.2004.00270.x

Van Wyk, R., Boschoff, A., & Cilliers, F. (2003). The prediction of job involvement for pharmacists and accountants. *South Africa Journal of Industrial Psychology, 29*(3), 61-67.

Van Wyk, R., Boshoff, A., & Bester, C. (2003). Entrepreneurial attitudes: What are their sources? *South African Journal of Economic and Management Sciences, 6*(1), 1-24.

vanDellen, M. R., Campbell, W. K., Hoyle, R. H., & Bradfield, E. K. (2010). Compensating, resisting, and breaking: A meta-analytic examination of reactions to self-esteem threat. *Personality and Social Psychology Review, 15*(1), 51-74. doi:10.1177/1088868310372950

Vazire, S., & Wilson, T. D. (Eds.). (2012). *Handbook of self-knowledge*. New York: Guilford Press.

Vecina, M. L., Chacón, F., & Pérez-Viejo, J. M. (2016). Moral absolutism, self-deception, and moral self-concept in men who commit intimate partner violence: A comparative study with an opposite sample. *Violence Against Women, 22*(1), 3-16. doi:10.1177/1077801215597791

Vinchur, A. J., & Koppes Bryan, L. L. (2012). A history of personnel assessment and selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 9-30). New York: Oxford University Press.

von Eye, A., & Nesselroade, J. R. (1992). Types of change: Application of configural frequency analysis in repeated measurement designs. *Experimental Aging Research, 18*, 169-183. doi:10.1080/03610739208260357

Voskuijl, O. (2005). Job analysis: Current and future perspectives. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 27-46). Oxford, England: Blackwell.

Weber, L. (2015, April 14). Today's personality tests raise the bar for job seekers. *The Wall Street Journal*. Retrieved from http://www.wsj.com/articles/a-personality-test-could-stand-in-the-way-of-your-next-job-1429065001

Whiteside-Mansell, L., & Corwyn, R. F. (2003). Mean and covariance structures analyses: An examination of the Rosenberg Self-Esteem Scale among adolescents and adults. *Educational and Psychological Measurement, 63*(1), 163-173. doi:10.1177/0013164402239323

Willett, J. B. (1988). Chapter 9: Questions and answers in the measurement of change. *Review of research in education, 15*(1), 345-422.

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49*, 587-602. doi:10.1177/001316448904900309

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381. doi:10.1037/0033-2909.116.2.363

Williams, R. H., & Zimmerman, D. W. (1996a). Are simple gain scores obsolete? *Applied Psychological Measurement, 20*(1), 59-69.

Williams, R. H., & Zimmerman, D. W. (1996b). Commentary on the commentaries of Collins and Humphreys. *Applied Psychological Measurement, 20*, 295-297. doi:10.1177/014662169602000310

Wilson, J. H., & Joye, S. W. (2017). *Research methods and statistics: An integrated approach*. Thousand Oaks, CA: Sage.

Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 505-560). New York: Plenum Press.

Wood, C., Conner, M., Sandberg, T., Godin, G., & Sheeran, P. (2014). Why does asking questions change health behaviours? The mediating role of attitude accessibility. *Psychology & Health, 29*(4), 390-404. doi:10.1080/08870446.2013.858343

Wylie, R. C. (1961). *The self concept: A critical survey of pertinent research literature*. Lincoln, NE: University of Nebraska Press.

Wylie, R. C. (1974). *The self-concept* (Rev. ed. Vol. 1). Lincoln, NE: University of Nebraska Press.

Wylie, R. C. (1979). *The self-concept* (Rev. ed. Vol. 2). Lincoln, NE: University of Nebraska Press.

Yancey, W. L., Rigsby, L., & McCarthy, J. D. (1972). Social position and self-evaluation: The relative importance of race. *American Journal of Sociology, 78*(2), 338-359.

Yanico, B. J., & Lu, T. G. C. (2000). A psychometric evaluation of the Six-Factor Self-Concept Scale in a sample of racial/ethnic minority women. *Educational and Psychological Measurement, 60*(1), 86-99. doi:10.1177/00131640021970385

Zimmerman, D., Ownsworth, T., O'Donovan, A., Roberts, J., & Gullo, M. J. (2017). Associations between executive functions and mental health outcomes for adults with autism spectrum disorder. *Psychiatry Research, 253*, 360-363. doi:10.1016/j.psychres.2017.04.023

Zimmerman, D. W., Brotohusodo, T. L., & Williams, R. H. (1981). The reliability of sums and differences of test scores: Some new results and anomalies. *Journal of Experimental Education, 49*, 177-186. doi:10.1080/00220973.1981.11011781

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement, 19*(2), 149-154. doi:10.1111/j.1745-3984.1982.tb00124.x

Appendix 2 – Preliminary survey

Summary of results

Uses personality assessments for selection (some or all advertised jobs)	Number of organisations
Yes	19
No	6

Personality assessment(s) used	Number of organisations
OPQ	9
15FQ+	8
Hogan	4
MBTI	1
DISC	1
TMI	1
Unspecified personality assessment(s)	4

Vacancies for which personality assessment is used	Number of organisations
All	3
Management/senior roles only	5
Various other specified or unspecified roles	8
At hiring manager's discretion	2

Approximate number of assessments per year	Number of organisations
0 – 50	8
50 – 100	2
100 - 150	3
150 – 200	0
More than 200	2
Unsure	4

Appendix 3 – Distribution of responses for researcher-created questionnaire items – Sub-project A

Frequency tables - Context of job application (before completion of personality assessment)

Are you currently employed?			
		Count	Percentage
Valid	Yes, in fulltime work	13	48.15
	Yes, in part-time work	4	14.81
	No	10	37.04
	Total	27	100.0

What is the main reason you are applying for a new job?			
		Count	Percentage
Valid	Currently/soon to be unemployed	9	33.33
	Looking for a job with higher pay	1	3.70
	Looking for a job with greater challenge/interest	10	37.04
	Change of career/occupation	4	14.81
	Other	3	11.11
	Total	27	100.0

How many roles have you applied for in the last month (including the role you are currently applying for)?			
		Count	Percentage
Valid	0	1	3.70
	1	4	14.81
	2	5	18.52
	3	6	22.22
	4	4	14.81
	5	1	3.70
	6	1	3.70
	7	1	3.70
	10	3	11.11
	50	1	3.70
	Total	27	100.0

What is the expected salary of the role you are applying for?			
		Count	Percentage
Valid	40,000 - 49,999	2	7.41
	50,000 - 59,999	3	11.11
	60,000 - 69,999	11	40.74
	70,000 - 79,999	5	18.52
	90,000 - 99,999	4	14.81
	110,000 - 119,999	1	3.70
	120,000+	1	3.70
	Total	27	100.0

Frequency tables - Expectations regarding psychometric assessments within the selection process (before completion of personality assessment)

Overall, do you expect completing the personality assessment to be a positive or negative experience for you?			
		Count	Percentage
Valid	Very positive	3	11.11
	Somewhat positive	11	40.74
	Neither positive nor negative	9	33.33
	Somewhat negative	4	14.81
	Total	27	100.0
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.			
		Count	Percentage
Valid	(25% confidence)	2	7.41
	(50% confidence)	12	44.44
	(75% confidence)	11	40.74
	Total confidence (100%)	2	7.41
	Total	27	100.0
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your psychometric ability assessment results.			
		Count	Percentage
Valid	(25% confidence)	1	3.70
	(50% confidence)	10	37.04
	(75% confidence)	12	44.44
	Total confidence (100%)	3	11.11
	Total	26	96.30
Missing	System	1	3.70
Total		27	100.0

Frequency tables - Previous experience of personality assessments (before completion of personality assessment)

Have you ever completed a personality assessment before? (I.e., filled out a formalised personality questionnaire or test)			
		Count	Percentage
Valid	Yes	17	62.96
	No	10	37.04
	Total	27	100.0

Please select the purpose(s) for which you have previously completed a personality assessment. Select all that apply.	Count
As part of a job application	11
Personal development (at work)	10
Personal development (outside of work)	6
As a student in school/university setting	5
Other (free-text response was, “Don’t remember”)	1

Which of the personality assessments below have you previously completed?	Count	Frequency
15FQ+	1	1
16 Personality Factor Questionnaire (16PF)	3	3
Clifton Strengths Finder (CSF)	0	0
DISC Assessment	3	3
Eysenck Personality Inventory (EPI)	1	1
Hogan Personality Inventory (HPI)	1	1
Millon Clinical Multiaxial Inventory (MCMI)	0	0
Minnesota Multiphasic Personality Inventory (MMPI)	3	3
Myers-Briggs Type Indicator (MBTI)	13	23
Occupational Personality Questionnaire (OPQ)	5	6
Revised NEO Personality Inventory (NEO-PI-R)	2	2
Rorschach Inkblot Test	0	0
Thematic Apperception Test (TAT)	0	0
Team Management Profile (TMP)	3	3

Frequency tables - Experience of completing personality assessment (after completion of personality assessment)

		Count
Overall, was completing the personality assessment a positive or negative experience for you?	Very positive	0
	Somewhat positive	6
	Neither positive nor negative	10
	Somewhat negative	9
	Very negative	2
Did you have any difficulty completing the personality assessment?	Yes	10
	No	17
Now that you have completed the personality assessment, select below from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results.	No confidence (0%)	1
	(25% confidence)	4
	(50% confidence)	12
	(75% confidence)	9
	Total confidence (100%)	1
Did the personality assessment allow you to express a true and complete picture of your personality?	Not at all	6
	A little	7
	Somewhat	13
	A lot	1
	Completely	0
Were there any questions in the personality assessment that did not apply to you as a person?	Yes	10
	No	17
While you were completing the personality assessment, did you think about the consistency of your responses, or try to make later responses consistent with earlier responses?	Yes, I thought about consistency and tried to make my responses consistent with each other	9

		Count
	Yes, I thought about consistency but it did not affect how I answered the questions	15
	No, I did not think about the consistency of my responses	3
Do you think the personality assessment was relevant to your job application?	Yes	20
	No	7

Frequency tables - Impact of completing personality assessment (after completion of personality assessment)

						Count
Were there any questions in the personality assessment that asked you something about yourself that you had never considered before?	Yes					5
	No					22
Did completing the personality assessment cause you to think differently about yourself?	Yes	Did this change in thinking about yourself make you feel better or worse about yourself?	Better	Was this change in thinking about yourself related to a specific area of your life, or did it apply more generally to you as a whole person?	Specific area or areas	2
					As a whole person	0
			Worse		Specific area or areas	1
					As a whole person	1
			No better or worse		Specific area or areas	1
					As a whole person	2
	No					20
Do you think if you had completed this personality assessment in the past, or if you took it in the future that you would answer some of the items differently?	Yes					20
	No					7
Did you find the personality assessment of any personal use?	Yes					6
	No					21
Did completing the personality assessment have any impact on your motivation to apply for other jobs?	Yes, it increased my motivation					1
	Yes, it decreased my motivation					4
	No					22
Did completing the personality assessment have	Yes, it increased my optimism					1

any impact on your optimism about successfully gaining the job you applied for, or future jobs?	Yes, it decreased my optimism	15
	No	11

Frequency tables - Outcome of job application (after completion of personality assessment)

		Count
Have you been notified of the outcome of your job application?	Yes	1
	No	26
What was the outcome of your job application?	Successful - I have been offered the job	1
	Ongoing - I have been shortlisted	0
	Unsuccessful - I have not been shortlisted or offered the job	0
	Other	0
Have you accepted the job offer?	Yes	0
	No, but I intend to accept	1
	No, I am still weighing my options	0
	No, but I intend to decline	0

Appendix 4 – Descriptive statistics for researcher-created questionnaire items – Sub-project A

Item-level descriptive statistics (interval-level variables; before completion of personality assessment)

	N	Range	Minimum	Maximum	$\bar{x}$	SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
How many roles have you applied for in the last month (including the role you are currently applying for)?	27	50	0	50	5.41	9.328	87.020	4.512	.448	21.955	.872
What is the expected salary of the role you are applying for?	27	8	3	11	5.78	1.928	3.718	1.140	.448	1.220	.872
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	27	3	1	4	2.52	.893	.798	.115	.448	-.600	.872
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.	27	3	2	5	3.48	.753	.567	.068	.448	-.107	.872
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your psychometric ability assessment results.	26	3	2	5	3.65	.745	.555	.052	.456	-.208	.887
Valid N	26										

Tests of Normality (before completion of personality assessment)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
How many roles have you applied for in the last month (including the role you are currently applying for)?	.321	26	.000	.411	26	.000
What is the expected salary of the role you are applying for?	.268	26	.000	.863	26	.003
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	.224	26	.002	.883	26	.007
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.	.244	26	.000	.858	26	.002
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your psychometric ability assessment results.	.256	26	.000	.850	26	.001
a. Lilliefors Significance Correction						

Item-level descriptive statistics (interval-level variables; after completion of personality assessment)

	N	Range	Minimum	Maximum	x̄	SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Overall, was completing the personality assessment a positive or negative experience for you?	27	3	2	5	3.26	.903	.815	.116	.448	-.737	.872
Now that you have completed the personality assessment, select below from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results.	27	4	1	5	3.19	.879	.772	-.387	.448	.321	.872
Did the personality assessment allow you to express a true and complete picture of your personality?	27	3	1	4	2.33	.877	.769	-.369	.448	-.978	.872
Valid N	27										

Tests of Normality (after completion of personality assessment)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall, was completing the personality assessment a positive or negative experience for you?	.206	27	.005	.877	27	.004
Now that you have completed the personality assessment, select below from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results.	.231	27	.001	.887	27	.007
Did the personality assessment allow you to express a true and complete picture of your personality?	.295	27	.000	.823	27	.000
a. Lilliefors Significance Correction						

Appendix 5 – Correlations between researcher-created questionnaire items – Sub-project A

Correlation matrix - Context of job application (before completion of personality assessment)

		How many roles have you applied for in the last month (including the role you are currently applying for)?	What is the expected salary of the role you are applying for?
How many roles have you applied for in the last month (including the role you are currently applying for)?	Pearson's r	1	.518**
	Sig. (2-tailed)		.006
	N	27	27
What is the expected salary of the role you are applying for?	Pearson's r	.518**	1
	Sig. (2-tailed)	.006	
	N	27	27
**. Correlation is significant at the .01 level			

Correlation matrix - Expectations regarding psychometric assessments (before completion of personality assessment)

		Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.	Please select below from 1 to 5 according to your confidence that / the hiring manager(s) will correctly interpret your psychometric ability assessment results.
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Pearson's r	1	-.271	-.365
	Sig. (2-tailed)		.171	.067
	N	27	27	26
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.	Pearson's r	-.271	1	.599**
	Sig. (2-tailed)	.171		.001
	N	27	27	26
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your psychometric ability assessment results.	Pearson's r	-.365	.599**	1
	Sig. (2-tailed)	.067	.001	
	N	26	26	26
**. Correlation is significant at the .01 level				

Correlation matrix - Experience of completing personality assessment (after completion of personality assessment)

		Overall, was completing the personality assessment a positive or negative experience for you?	Now that you have completed the personality assessment, select below from 1 to 5 according to your confidence that the hiring manager will correctly interpret your results.	Did the personality assessment allow you to express a true and complete picture of your personality?
Overall, was completing the personality assessment a positive or negative experience for you?	Pearson's r	1	-.596**	-.356
	Sig. (2-tailed)		.001	.068
	N	27	27	27
Now that you have completed the personality assessment, select below from 1 to 5 according to your confidence that the hiring manager will correctly interpret your personality assessment results.	Pearson's r	-.596**	1	.616**
	Sig. (2-tailed)	.001		.001
	N	27	27	27
Did the personality assessment allow you to express a true and complete picture of your personality?	Pearson's r	-.356	.616**	1
	Sig. (2-tailed)	.068	.001	
	N	27	27	27
**. Correlation is significant at the .01 level				

Appendix 6 - Reliability analysis for researcher-created questionnaire items – Sub-project A

Reliability statistics – Expectations regarding psychometric assessments (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.669	.683	3

Item-Total statistics – Expectations regarding psychometric assessments (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's α if item deleted
Confidence in interpretation of personality assessment results.	7.12	1.866	.519	.365	.528
Confidence in interpretation of psychometric ability assessment results.	6.96	1.798	.586	.399	.445
Expectation of assessment as positive or negative experience (reverse scored)	7.15	1.815	.366	.141	.749

Reliability statistics – Experience of completing personality assessment (after completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.766	.767	3

Item-Total statistics – Experience of completing personality assessment (after completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's α if item deleted
Confidence in interpretation of personality assessment results.	5.07	2.148	.736	.542	.525
Assessment allowed true and complete picture of your personality	5.93	2.533	.542	.379	.747
Experience of assessment as positive or negative experience (reverse scored)	5.52	2.490	.530	.356	.762

Appendix 7 - Descriptive statistics for SFSCS subscales and composite score – Sub-project A

Before completion of personality assessment

Sub-project A - before completion of personality assessment	N	Range	Minimum	Maximum	$\bar{x}$		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Likeability	27	13	29	42	35.63	.738	3.835	14.704	.351	.448	-.941	.872
Task Accomplishment	27	15	27	42	36.07	.720	3.741	13.994	-.624	.448	-.111	.872
Power (weighted)	27	16.29	18.86	35.14	24.540	.900	4.678	21.880	.563	.448	-.693	.872
Vulnerability	27	17	12	29	20.56	.951	4.941	24.410	-.012	.448	-1.039	.872
Gifted (weighted)	27	25.20	15.60	40.80	29.244	1.357	7.054	49.752	-.170	.448	-.748	.872
Moral	27	11	31	42	38.70	.582	3.023	9.140	-1.064	.448	.711	.872
Total self-concept	27	55.29	113.51	168.80	143.636	3.313	17.215	296.354	-.284	.448	-1.120	.872
Valid N	27											

After completion of personality assessment

Sub-project A - after completion of personality assessment	N	Range	Minimum	Maximum	x̄		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Likeability	27	15	27	42	35.56	.805	4.182	17.487	-.264	.448	-.525	.872
Task Accomplishment	27	22	20	42	35.74	.754	3.918	15.353	-2.471	.448	9.803	.872
Power (weighted)	27	18.00	13.71	31.71	23.016	.970	5.041	25.415	.206	.448	-.897	.872
Vulnerability	27	19	10	29	19.44	1.103	5.733	32.872	.145	.448	-1.293	.872
Gifted (weighted)	27	28.80	12.00	40.80	28.756	1.461	7.593	57.653	-.326	.448	-.426	.872
Moral	27	13	29	42	38.59	.630	3.273	10.712	-1.356	.448	1.774	.872
Total self-concept	27	79.51	90.86	170.37	142.216	3.812	19.810	392.416	-.757	.448	.006	.872
Valid N	27											

Appendix 8 - Reliability analysis for SFSCS – Sub-project A

Reliability statistics for SFSCS (all items; before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.878	.883	36

Item-Total statistics for SFSCS (all items; before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Fun to be with	185.111	270.410	.452	.875
Hard worker	184.778	277.256	.160	.879
Dominant	186.630	263.242	.509	.873
Easily embarrassed	186.000	268.923	.286	.878
A natural talent	186.296	251.909	.603	.870
Loyal	184.333	273.615	.305	.877
Strong	185.444	263.718	.654	.871
Friendly	184.593	279.097	.147	.879
Productive	184.704	273.755	.293	.877
Lacks confidence	186.000	260.154	.409	.875
Law-abiding	184.444	276.256	.223	.878
Forceful	187.482	273.875	.152	.881
Has special talents	186.222	251.179	.679	.868
Plans ahead	185.259	272.353	.298	.877
Sociable	185.259	267.430	.404	.875
Easily hurt	186.296	272.447	.184	.880
Acts as a leader	185.630	258.627	.624	.870

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Truthful	184.593	273.558	.386	.876
Self-conscious	187.333	281.231	-.020	.886
Works efficiently	184.889	268.333	.541	.873
Faithful	184.407	272.405	.411	.875
Aggressive	188.519	275.028	.241	.878
Easy to talk to	185.148	264.746	.474	.873
Bright and ingenious	185.630	254.858	.650	.869
Makes mistakes when flustered	186.074	262.687	.460	.874
Honest	184.482	275.259	.321	.877
Good at meeting deadlines	184.778	267.949	.378	.875
Pleasant	184.593	280.097	.084	.879
Powerful	186.630	257.550	.564	.871
Creative	185.8519	250.670	.623	.869
Easily rattled when people are watching	185.963	261.191	.550	.872
Trustworthy	184.148	274.285	.434	.876
Can concentrate well on a task	184.630	274.088	.337	.876
Warm	184.778	274.872	.248	.878
Tough	187.000	263.462	.362	.876
Has innate ability	185.889	255.026	.555	.871

Reliability statistics for SFSCS subscale Likeability (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.850	.853	6

Item-Total statistics for SFSCS subscale Likeability (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Fun to be with	29.89	10.641	.698	.815
Friendly	29.37	12.088	.547	.843
Sociable	30.04	9.652	.608	.836
Easy to talk to	29.93	8.379	.839	.780
Pleasant	29.37	11.858	.535	.843
Warm	29.56	10.487	.658	.820

Reliability statistics for SFSCS subscale Task Accomplishment (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.820	.819	6

Item-Total statistics for SFSCS subscale Task Accomplishment (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Hard worker	30.00	10.154	.597	.789
Productive	29.93	9.687	.714	.764
Plans ahead	30.48	9.567	.619	.784
Works efficiently	30.11	10.256	.651	.780
Good at meeting deadlines	30.00	8.769	.642	.782
Can concentrate well on a task	29.85	11.977	.317	.838

Reliability statistics for SFSCS subscale Power (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.802	.806	7

Item-Total statistics for SFSCS subscale Power (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Dominant	24.41	22.712	.569	.770
Strong	23.22	24.641	.528	.781
Forceful	25.26	21.815	.496	.785
Acts as a leader	23.41	23.943	.418	.796
Aggressive	26.30	24.909	.503	.785
Powerful	24.41	20.943	.614	.760
Tough	24.78	19.256	.673	.747

Reliability statistics for SFSCS subscale Vulnerability (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.683	.705	6

Item-Total statistics for SFSCS subscale Vulnerability (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Easily embarrassed	17.41	17.097	.548	.597
Lacks confidence	17.41	18.251	.300	.686
Easily hurt	17.11	17.103	.498	.612
Self-conscious	16.07	20.840	.111	.745
Makes mistakes when flustered	17.33	17.846	.493	.617
Easily rattled when people are watching	17.44	17.026	.662	.570

Reliability statistics for SFSCS subscale Gifted (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.893	.896	5

Item-Total statistics for SFSCS subscale Gifted (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
A natural talent	19.81	22.080	.742	.869
Has special talents	19.74	21.892	.847	.845
Bright and ingenious	19.15	23.516	.781	.862
Creative	19.37	23.550	.605	.901
Has innate ability	19.41	22.481	.741	.869

Reliability statistics for SFSCS subscale Moral (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.831	.842	6

Item-Total statistics for SFSCS subscale Moral (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's α if item deleted
Loyal	32.19	6.464	.498	.831
Law-abiding	32.30	6.832	.448	.838
Truthful	32.44	6.179	.776	.769
Faithful	32.26	6.046	.759	.770
Honest	32.33	6.923	.550	.814
Trustworthy	32.00	6.923	.675	.796

Appendix 9 - Correlations between self-concept variables (SFSCS subscale scores) and self-esteem (RSES total score) – Sub-project A

Before completion of personality assessment

Sub-project A - before completion of personality assessment		Likeability	Task Accomplishment	Power (weighted)	Vulnerability	Gifted (weighted)	Moral	Self-esteem
SELF-CONCEPT Likeability	Pearson's r	1	.061	-.018	-.249	.354	.259	.316
	Sig. (2-tailed)		.763	.93	.211	.07	.192	.109
	N	27	27	27	27	27	27	27
Task Accomplishment	Pearson's r	.061	1	.29	-.306	.167	.206	.456*
	Sig. (2-tailed)	.763		.143	.12	.406	.302	.017
	N	27	27	27	27	27	27	27
Power (weighted)	Pearson's r	-.018	.29	1	-.262	.537**	.338	.520**
	Sig. (2-tailed)	.93	.143		.187	.004	.085	.005
	N	27	27	27	27	27	27	27
Vulnerability	Pearson's r	-.249	-.306	-.262	1	-.293	.089	-.629**
	Sig. (2-tailed)	.211	.12	.187		.137	.66	0
	N	27	27	27	27	27	27	27
Gifted (weighted)	Pearson's r	.354	.167	.537**	-.293	1	.398*	.551**
	Sig. (2-tailed)	.07	.406	.004	.137		.04	.003
	N	27	27	27	27	27	27	27
Moral	Pearson's r	.259	.206	.338	.089	.398*	1	.218
	Sig. (2-tailed)	.192	.302	.085	.66	.04		.274
	N	27	27	27	27	27	27	27
SELF-ESTEEM Self-esteem	Pearson's r	.316	.456*	.520**	-.629**	.551**	.218	1
	Sig. (2-tailed)	.109	.017	.005	0	.003	.274	
	N	27	27	27	27	27	27	27
*. Correlation is significant at the .05 level								

**. Correlation is significant at the .01 level

After completion of personality assessment

Sub-project A - after completion of personality assessment		Likeability	Task Accomplishment	Power (weighted)	Vulnerability	Gifted (weighted)	Moral	Self-esteem
SELF-CONCEPT Likeability	Pearson's r	1	.427*	.141	-.397*	.337	.264	.547**
	Sig. (2-tailed)		.026	.483	.04	.086	.183	.003
	N	27	27	27	27	27	27	27
Task Accomplishment	Pearson's r	.427*	1	.157	-.443*	.426*	.378	.740**
	Sig. (2-tailed)	.026		.435	.021	.027	.052	0
	N	27	27	27	27	27	27	27
Power (weighted)	Pearson's r	.141	.157	1	-.16	.450*	.161	.255
	Sig. (2-tailed)	.483	.435		.426	.018	.424	.199
	N	27	27	27	27	27	27	27
Vulnerability	Pearson's r	-.397*	-.443*	-.16	1	-.375	-.076	-.719**
	Sig. (2-tailed)	.04	.021	.426		.054	.706	0
	N	27	27	27	27	27	27	27
Gifted (weighted)	Pearson's r	.337	.426*	.450*	-.375	1	.395*	.543**
	Sig. (2-tailed)	.086	.027	.018	.054		.042	.003
	N	27	27	27	27	27	27	27
Moral	Pearson's r	.264	.378	.161	-.076	.395*	1	.394*
	Sig. (2-tailed)	.183	.052	.424	.706	.042		.042
	N	27	27	27	27	27	27	27
SELF-ESTEEM Self-esteem	Pearson's r	.547**	.740**	.255	-.719**	.543**	.394*	1
	Sig. (2-tailed)	.003	0	.199	0	.003	.042	
	N	27	27	27	27	27	27	27
*. Correlation is significant at the .05 level								
**. Correlation is significant at the .01 level								

Appendix 10 - Descriptive statistics for self-esteem (RSES total score) – Sub-project A

Before and after completion of personality assessment

	N	Range	Minimum	Maximum	$\bar{x}$		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Before completion of personality assessment	27	21	9	30	22.70	.960	4.991	24.909	-.654	.448	.695	.872
After completion of personality assessment	27	25	5	30	22.74	1.109	5.762	33.199	-1.065	.448	1.939	.872
Valid N	27											

Appendix 11 – Reliability analysis for RSES – Sub-project A

Reliability statistics for RSES (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.877	.879	10

Item-Total statistics for RSES (before completion of personality assessment)

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation	Cronbach's α if item deleted
I feel that I am a person of worth, at last on an equal plane with others	20.19	21.695	.623	.812	.867
I feel that I have a number of good qualities	20.04	22.729	.425	.695	.877
All in all, I am inclined to feel that I am a failure	20.30	20.524	.621	.628	.864
I am able to do things as well as most other people	20.19	22.311	.413	.613	.878
I feel I do not have much to be proud of	20.52	21.259	.384	.388	.884
I take a positive attitude toward myself	20.37	20.011	.730	.754	.856
On the whole, I am satisfied with myself	20.52	19.567	.738	.813	.855
I wish I could have more respect for myself	20.59	20.097	.581	.598	.867
I certainly feel useless at times	21.00	18.154	.762	.684	.852
At times I think I am no good at all	20.63	18.165	.803	.805	.848

Appendix 12 – Component structure for RSES – Sub-project A

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.702
Bartlett's Test of Sphericity	Approx. χ^2	150.864
	df	45
	Sig.	.000

Component structure and proportion of variance explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.944	49.438	49.438	4.944	49.438	49.438	3.849	38.490	38.490
2	1.451	14.509	63.947	1.451	14.509	63.947	2.145	21.450	59.939
3	1.041	10.407	74.354	1.041	10.407	74.354	1.441	14.414	74.354
Extraction Method: Principal Component Analysis.									

Appendix 13 – Descriptive statistics for self-concept and self-esteem change scores – Sub-project A

	N	Range	Minimum	Maximum	$\bar{x}$		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SELF-CONCEPT Change in Likeability	27	10.00	-7.00	3.00	-.074	.462	2.401	5.764	-1.275	.448	1.655	.872
Change in Task Accomplishment	27	12.00	-7.00	5.00	-.333	.528	2.746	7.538	-.321	.448	.282	.872
Change in Power	27	9.43	-5.14	4.29	-1.524	.456	2.369	5.614	.235	.448	.145	.872
Change in Vulnerability	27	11.00	-6.00	5.00	-1.111	.553	2.873	8.256	.120	.448	-.333	.872
Change in Gifted	27	9.60	-4.80	4.80	-.489	.556	2.887	8.336	.289	.448	-.734	.872
Change in Moral	27	10.00	-6.00	4.00	-.111	.355	1.847	3.410	-.893	.448	3.263	.872
Change in Total self-concept	27	32.63	-23.46	9.17	-1.420	1.519	7.895	62.333	-1.446	.448	2.726	.872
SELF-ESTEEM Change in Self-esteem	27	8.00	-4.00	4.00	.037	.415	2.157	4.652	.222	.448	-.254	.872
Valid N	27											

Appendix 14 - Normality of distributions for self-concept and self-esteem change scores - Sub-project A

Kolmogorov-Smirnov test

	Statistic	df	Sig.
Change in Likeability	.191	27	.012
Change in Task Accomplishment	.145	27	.153
Change in Power	.135	27	.200*
Change in Vulnerability	.132	27	.200*
Change in Gifted	.116	27	.200*
Change in Moral	.163	27	.065
Change in Total self-concept	.191	27	.013
Change in Self-esteem	.142	27	.17
* This is a lower bound of the true significance			

Appendix 15 – Paired samples t-tests for self-concept and self-esteem scores – Sub-project A

	Paired Differences					t	df	Sig. (2-tailed)
	x	SD	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Likeability 1 versus Likeability 2	.074	2.401	.462	-.876	1.024	.160	26	.874
Task Accomplishment 1 versus Task Accomplishment 2	.333	2.746	.528	-.753	1.419	.631	26	.534
Power 1 (weighted) versus Power 2 (weighted)	1.524	2.369	.456	.587	2.461	3.342	26	.003
Vulnerability 1 versus Vulnerability 2	1.111	2.873	.553	-.026	2.248	2.009	26	.055
Gifted 1 (weighted) versus Gifted 2 (weighted)	.4889	2.887	.556	-.653	1.631	.880	26	.387
Moral 1 versus Moral 2	.111	1.847	.355	-.619	.842	.313	26	.757
Total Self-concept 1 (weighted) versus Total Self-concept 2 (weighted)	1.420	7.895	1.519	-1.703	4.543	.935	26	.359
Self-esteem 1 versus Self-esteem 2	-.037	2.157	.415	-.890	.816	-.089	26	.930

Appendix 16 – Analysis of variance for potential predictors of self-perception change scores (MANOVA) – Sub-project A

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.845	1.824[a]	3.000	1.000	.487
	Wilks' Lambda	.155	1.824[a]	3.000	1.000	.487
	Hotelling's Trace	5.472	1.824[a]	3.000	1.000	.487
	Roy's Largest Root	5.472	1.824[a]	3.000	1.000	.487
Are you currently employed?	Pillai's Trace	1.171	.942	6.000	4.000	.549
	Wilks' Lambda	.119	.631[a]	6.000	2.000	.720
	Hotelling's Trace	4.938	.000	6.000	.000	.
	Roy's Largest Root	4.382	2.922[b]	3.000	2.000	.265
What is the main reason you are applying for a new job?	Pillai's Trace	1.664	.935	12.000	9.000	.554
	Wilks' Lambda	.017	.889	12.000	2.937	.623
	Hotelling's Trace	.	.	12.000	.	.
	Roy's Largest Root	15.489	11.617[b]	4.000	3.000	.036
What is the expected salary of the role you are applying for?	Pillai's Trace	1.960	1.131	15.000	9.000	.440
	Wilks' Lambda	.025	.591	15.000	3.162	.790
	Hotelling's Trace	.	.	15.000	.	.
	Roy's Largest Root	7.395	4.437[b]	5.000	3.000	.125
Are you currently employed? * What is the main reason you are applying for a new job?	Pillai's Trace	.695	.759[a]	3.000	1.000	.666
	Wilks' Lambda	.305	.759[a]	3.000	1.000	.666
	Hotelling's Trace	2.276	.759[a]	3.000	1.000	.666
	Roy's Largest Root	2.276	.759[a]	3.000	1.000	.666
Are you currently employed?	Pillai's Trace	1.228	1.061	6.000	4.000	.500

* What is the expected salary of the role you are applying for?	Wilks' Lambda	.051	1.143[a]	6.000	2.000	.536
	Hotelling's Trace	13.134	.000	6.000	.000	.
	Roy's Largest Root	12.703	8.468[b]	3.000	2.000	.107
What is the main reason you are applying for a new job? * What is the expected salary of the role you are applying for?	Pillai's Trace	1.414	.891	9.000	9.000	.567
	Wilks' Lambda	.029	.949	9.000	2.584	.594
	Hotelling's Trace	.	.	9.000	.	.
	Roy's Largest Root	17.652	17.652[b]	3.000	3.000	.021
Are you currently employed? * What is the main reason you are applying for a new job? * What is the expected salary of the role you are applying for?	Pillai's Trace	.000	.[a]	.000	.000	.
	Wilks' Lambda	1.000	.[a]	.000	2.000	.
	Hotelling's Trace	.000	.[a]	.000	2.000	.
	Roy's Largest Root	.000	.000[a]	3.000	.000	.
a. Exact statistic						
b. The statistic is an upper bound on F that yields a lower bound on the significance level.						

Appendix 17 - Correlations between self-concept and self-esteem change scores - Sub-project A

Sub-project A – change scores		Change in Likeability	Change in Task Accomplishment	Change in Power	Change in Vulnerability	Change in Gifted	Change in Moral	Change in Self-esteem
SELF-CONCEPT Change in Likeability	Pearson's r	1	.352	.072	-.453*	.048	.033	.327
	Sig. (2-tailed)		.072	.721	.018	.813	.871	.096
	N	27	27	27	27	27	27	27
Change in Task Accomplishment	Pearson's r	.352	1	.091	-.161	.124	.357	.171
	Sig. (2-tailed)	.072		.651	.423	.537	.068	.394
	N	27	27	27	27	27	27	27
Change in Power (weighted)	Pearson's r	.072	.091	1	.178	.280	-.018	-.034
	Sig. (2-tailed)	.721	.651		.376	.157	.931	.868
	N	27	27	27	27	27	27	27
Change in Vulnerability	Pearson's r	-.453*	-.161	.178	1	-.085	.114	-.564**
	Sig. (2-tailed)	.018	.423	.376		.675	.573	.002
	N	27	27	27	27	27	27	27
Change in Gifted (weighted)	Pearson's r	.048	.124	.280	-.085	1	-.002	-.019
	Sig. (2-tailed)	.813	.537	.157	.675		.992	.924
	N	27	27	27	27	27	27	27
Change in Moral	Pearson's r	.033	.357	-.018	.114	-.002	1	.069
	Sig. (2-tailed)	.871	.068	.931	.573	.992		.734
	N	27	27	27	27	27	27	27
SELF-ESTEEM Change in Self-esteem	Pearson's r	.327	.171	-.034	-.564**	-.019	.069	1
	Sig. (2-tailed)	.096	.394	.868	.002	.924	.734	
	N	27	27	27	27	27	27	27
*. Correlation is significant at the .05 level								
**. Correlation is significant at the .01 level								

Appendix 18 – Distribution of responses for researcher-created questionnaire items – Sub-project B

Frequency tables - Expectations regarding personality assessment (before completion of personality assessment)

Overall, do you expect completing the personality assessment to be a positive or negative experience for you?			
		Count	Percentage
Valid	Very positive	14	15.22
	Somewhat positive	61	66.30
	Neither positive nor negative	15	16.30
	Somewhat negative	2	2.17
	Total	92	100.0
Please select below from 1 to 5 according to your confidence that the hiring manager(s) will correctly interpret your personality assessment results.			
		Count	Percentage
Valid	(25% confidence)	1	1.09
	(50% confidence)	26	28.26
	(75% confidence)	57	61.96
	Total confidence (100%)	8	8.70
	Total	92	100.0

Frequency tables - Previous experience of personality assessments (before completion of personality assessment)

Have you ever completed a personality assessment before? (I.e., filled out a formalised personality questionnaire or test)			
		Count	Percentage
Valid	Yes	44	47.83
	No	48	52.17
	Total	92	100.0

Please select the purpose(s) for which you have previously completed a personality assessment. Select all that apply.	Count
As part of a job application	17
Personal development (at work)	15
Personal development (outside of work)	17
As a student in school/university setting	19

Which of the personality assessments below have you previously completed?	Count	Frequency
15FQ+	2	4
16 Personality Factor Questionnaire (16PF)	11	13
Clifton Strengths Finder (CSF)	6	10
DISC Assessment	5	6
Eysenck Personality Inventory (EPI)	1	1
Hogan Personality Inventory (HPI)	3	3
Millon Clinical Multiaxial Inventory (MCMI)	1	1
Minnesota Multiphasic Personality Inventory (MMPI)	4	4
Myers-Briggs Type Indicator (MBTI)	33	80
Occupational Personality Questionnaire (OPQ)	10	10
Revised NEO Personality Inventory (NEO-PI-R)	4	4
Rorschach Inkblot Test	1	1
Thematic Apperception Test (TAT)	2	3
Team Management Profile (TMP)	2	4

Frequency tables - Experience of completing personality assessment (after completion of personality assessment)

		Count
Overall, was completing the personality assessment a positive or negative experience for you?	Very positive	9
	Somewhat positive	36
	Neither positive nor negative	29
	Somewhat negative	18
	Very negative	0
Did you have any difficulty completing the personality assessment?	Yes	30
	No	62
Did the personality assessment allow you to express a true and complete picture of your personality?	Not at all	7
	A little	17
	Somewhat	49
	A lot	10
	Completely	9
Were there any questions in the personality assessment that did not apply to you as a person?	Yes	40
	No	52
While you were completing the personality assessment, did you think about the consistency of your responses, or try to make later responses consistent with earlier responses?	Yes, I thought about consistency and tried to make my responses consistent with each other	27
	Yes, I thought about consistency but it did not affect how I answered the questions	53
	No, I did not think about the consistency of my responses	12

Frequency tables - Impact of completing personality assessment (after completion of personality assessment)

						Count
Were there any questions in the personality assessment that asked you something about yourself that you had never considered before?	Yes					22
	No					70
Did completing the personality assessment cause you to think differently about yourself?	Yes	Did this change in thinking about yourself make you feel better or worse about yourself?	Better	Was this change in thinking about yourself related to a specific area of your life, or did it apply more generally to you as a whole person?	Specific area or areas	4
					As a whole person	2
			Worse		Specific area or areas	2
					As a whole person	2
			No better or worse		Specific area or areas	7
					As a whole person	5
	No					0
Did completing the personality assessment have any impact on your optimism about successfully gaining future jobs?	Yes, it increased my optimism					15
	Yes, it decreased my optimism					8
	No					69

Frequency tables - Experience of receiving feedback (after completion of personality assessment)

Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?			
		Frequency	Percentage
Valid	No detail	1	1.09
	Very little detail	5	5.43
	Some detail	39	42.39
	A lot of detail	41	44.57
	Full detail	6	6.52
	Total	92	100.0

Was the written feedback from your personality assessment clear and easy to understand?			
		Frequency	Percentage
Valid	I didn't understand it at all	1	1.09
	I understood most of it	10	10.87
	I understood all of it	81	88.04
	Total	92	100.0

Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.			
		Frequency	Percentage
Valid	No confidence (0%)	1	1.09
	(25% confidence)	7	7.61
	(50% confidence)	25	27.17
	(75% confidence)	43	46.74
	Total confidence (100%)	16	17.39
	Total	92	100.0

Was receiving the written feedback from your personality assessment a positive or negative experience for you?			
		Frequency	Percentage
Valid	Very positive	14	15.22
	Somewhat positive	36	39.13
	Neither positive nor negative	35	38.04
	Somewhat negative	6	6.52
	Very negative	1	1.09
	Total	92	100.0

Did the written feedback from the personality assessment express a true and complete picture of your personality?			
		Frequency	Percentage
Valid	Yes	31	33.70
	No	61	66.30
	Total	92	100.0

Did any part of the feedback surprise you?			
		Frequency	Percentage
Valid	Yes	29	31.52
	No	63	68.48
	Total	92	100.0

Frequency tables - Impact of receiving feedback (after completion of personality assessment)

Did any part of the written feedback cause you to think something about yourself that you had never considered before?			
		Frequency	Percentage
Valid	Yes	26	28.26
	No	66	71.74
	Total	92	100.0

Did receiving the written feedback cause you to think differently about yourself?					Count
Yes	Did this change in thinking about yourself make you feel better or worse about yourself?	Better	Q4.19 Was this change in thinking about yourself related to a specific area of your life, or did it apply more generally to you as a whole / person?	Specific area	1
				As a whole person	3
		Worse		Specific area	2
				As a whole person	2
		No better or worse		Specific area	2
				As a whole person	3
No					0

Now that you have received the written feedback from the personality assessment, do you think if you had completed this personality assessment in the past, or if you completed it in the future that you would answer some of the items differently?			
		Frequency	Percentage
Valid	Yes	49	53.26
	No	43	46.74
	Total	92	100.0

Did you find the written feedback of any personal use?			
		Frequency	Percentage
Valid	Yes	48	52.17
	No	44	47.83
	Total	92	100.0

Did receiving the written feedback have any impact on your optimism about future jobs?			
		Frequency	Percentage
Valid	Yes, it increased my optimism	15	16.30
	Yes, it decreased my optimism	8	8.70
	No	69	75.00
	Total	92	100.0

Appendix 19 – Descriptive statistics for researcher-created questionnaire items – Sub-project B

Item-level descriptive statistics - Interval-level variables (before completion of personality assessment)

	N	Range		Minimum	Maximum	x̄		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic		Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	92	3		1	4	2.05	.066	.635	.404	.482	.251	1.075	.498
Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.	92	3		2	5	3.78	.063	.608	.370	-.155	.251	.122	.498
Valid N	92												

Tests of Normality (before completion of personality assessment)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	.349	92	.000	.771	92	.000
Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.	.346	92	.000	.773	92	.000
a. Lilliefors Significance Correction						

Item-level descriptive statistics - Interval-level variables (after completion of personality assessment)

	N	Range	Minimum	Maximum	x̄		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Overall, was completing the personality assessment a positive or negative experience for you?	92	3	1	4	2.61	.095	.913	.834	.068	.251	-.852	.498
Did the personality assessment allow you to express a true and complete picture of your personality?	92	4	1	5	2.97	.104	.999	.999	.201	.251	.248	.498
Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?	92	4	1	5	3.50	.078	.749	.560	-.321	.251	.626	.498
Was the written feedback from your personality assessment clear and easy to understand?	92	3	1	4	3.86	.045	.434	.189	-4.041	.251	20.741	.498
Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.	92	4	1	5	3.72	.092	.881	.776	-.493	.251	.124	.498
Was receiving the written feedback from your personality assessment a positive or negative experience?	92	4	1	5	2.39	.090	.864	.746	.188	.251	-.054	.498

Valid N	92											

Tests of Normality (after completion of personality assessment)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall, was completing the personality assessment a positive or negative experience for you?	.237	92	.000	.871	92	.000
Did the personality assessment allow you to express a true and complete picture of your personality?	.280	92	.000	.868	92	.000
Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?	.259	92	.000	.841	92	.000
Was the written feedback from your personality assessment clear and easy to understand?	.508	92	.000	.357	92	.000
Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.	.267	92	.000	.871	92	.000
Was receiving the written feedback from your personality assessment a positive or negative experience?	.218	92	.000	.877	92	.000
a. Lilliefors Significance Correction						

Appendix 20 – Correlations between researcher-created questionnaire items – Sub-project B

Correlation matrix - Expectations regarding personality assessment (before completion of personality assessment)

		Q3.1 Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Q3.3 Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.
Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Pearson's r	1	-.140
	Sig. (2-tailed)		.184
	N	92	92
Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.	Pearson's r	-.140	1
	Sig. (2-tailed)	.184	
	N	92	92

Correlation matrix - Experience of completing personality assessment (after completion of personality assessment)

		Overall, was completing the personality assessment a positive or negative experience for you?	Did the personality assessment allow you to express a true and complete picture of your personality?
Overall, was completing the personality assessment a positive or negative experience for you?	Pearson's r	1	-.532**
	Sig. (2-tailed)		.000
	N	92	92
Did the personality assessment allow you to express a true and complete picture of your personality?	Pearson's r	-.532**	1
	Sig. (2-tailed)	.000	
	N	92	92
**. Correlation is significant at the .01 level			

Correlation matrix - Experience of receiving written feedback (after completion of personality assessment)

		Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?	Was the written feedback from your personality assessment clear and easy to understand?	Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.	Was receiving the written feedback from your personality assessment a positive or negative experience?
Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?	Pearson's r	1	-.017	.350**	-.493**
	Sig. (2-tailed)		.873	.001	.000
	N	92	92	92	92
Was the written feedback from your personality assessment clear and easy to understand?	Pearson's r	-.017	1	.297**	-.056
	Sig. (2-tailed)	.873		.004	.596
	N	92	92	92	92
Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.	Pearson's r	.350**	.297**	1	-.402**
	Sig. (2-tailed)	.001	.004		.000
	N	92	92	92	92
Was receiving the written feedback from your personality assessment a positive or negative experience?	Pearson's r	-.493**	-.056	-.402**	1
	Sig. (2-tailed)	.000	.596	.000	
	N	92	92	92	92
**. Correlation is significant at the .01 level					

Appendix 21 - Reliability analysis for researcher-created questionnaire items – Sub-project B

Reliability statistics – Expectations regarding personality assessment (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.245	.245	2

Reliability statistics – Experience of completing personality assessment (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.692	.694	2

Reliability statistics – Experience of receiving feedback (before completion of personality assessment)

Cronbach's α	Cronbach's α based on standardised items	N of items
.619	.588	4

Item-Total statistics – Experience of receiving feedback (before completion of personality assessment)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's α if Item Deleted
Thinking about the written feedback you received regarding your personality assessment: How would you rate the level of detail provided to you?	11.18	2.592	.451	.280	.510
Was the written feedback from your personality assessment clear and easy to understand?	10.83	3.794	.153	.105	.676
Now that you have received the written feedback from the personality assessment, select below from 1 to 5 according to your confidence that your personality assessment results were interpreted correctly.	10.97	2.164	.502	.272	.463
Was receiving the written feedback from your personality assessment a positive or negative experience? (reverse scored)	11.08	2.203	.503	.303	.461

Appendix 22 - Descriptive statistics for self-concept subscales and composite (SFSCS) score – Sub-project B

Before completion of personality assessment – Sub-project B

	N	Range	Minimum	Maximum	x̄	SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Likeability	92	23	19	42	32.45	5.436	29.546	-.793	.251	-.007	.498
Task Accomplishment	92	27	15	42	33.40	5.314	28.243	-1.106	.251	1.444	.498
Power (weighted)	92	24.86	9.43	34.29	23.441	5.348	28.604	.166	.251	-.532	.498
Vulnerability	92	33	7	40	25.27	6.766	45.782	-.135	.251	-.216	.498
Gifted (weighted)	92	32.40	9.60	42.00	27.274	6.051	36.620	-.077	.251	.677	.498
Moral	92	12	30	42	37.63	3.004	9.027	-.614	.251	-.217	.498
Total self-concept (weighted)	92	104.54	70.03	174.57	128.921	20.103	404.112	-.204	.251	.627	.498
Valid N	92										

After completion of personality assessment – Sub-project B

	N	Range	Minimum	Maximum	x̄	SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Likeability	92	28	14	42	33.05	5.763	33.217	-1.103	.251	1.080	.498
Task Accomplishment	92	28	14	42	33.43	5.364	28.776	-1.039	.251	1.297	.498
Power (weighted)	92	24.86	12.00	36.86	23.450	6.016	36.195	.316	.251	-.831	.498
Vulnerability	92	30	11	41	24.53	6.979	48.713	.018	.251	-.320	.498
Gifted (weighted)	92	30.00	12.00	42.00	27.835	6.605	43.619	-.094	.251	-.171	.498
Moral	92	14	28	42	37.32	3.288	10.812	-.445	.251	-.306	.498
Total self-concept (weighted)	92	111.14	69.00	180.14	130.557	21.587	465.988	-.111	.251	.145	.498
Valid N	92										

Appendix 23 - Reliability analysis for SFSCS – Sub-project B

Reliability statistics for SFSCS (entire test): Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.890	.889	36

Item-Total statistics for SFSCS (all items): Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Fun to be with	171.26	383.953	.318	.888
Hard worker	170.53	379.351	.483	.886
Dominant	172.51	376.538	.438	.886
A natural talent	172.11	375.285	.449	.886
Loyal	169.95	389.481	.281	.889
Strong	170.96	377.844	.496	.885
Friendly	170.50	384.626	.357	.887
Productive	170.87	373.653	.576	.884
Law-abiding	170.14	395.134	.110	.891
Forceful	173.29	383.858	.296	.889
Has special talents	172.26	370.766	.491	.885
Plans ahead	170.91	381.641	.300	.889
Sociable	171.71	378.495	.324	.889
Acts as a leader	171.95	364.601	.638	.882
Truthful	170.17	395.530	.150	.890
Works efficiently	170.75	376.651	.513	.885
Faithful	169.98	385.692	.378	.887

Aggressive	173.59	395.740	.050	.893
Easy to talk to	170.63	377.488	.445	.886
Bright and ingenious	171.23	375.936	.531	.885
Honest	170.03	395.043	.168	.890
Good at meeting deadlines	170.49	383.637	.301	.889
Pleasant	170.43	387.040	.306	.888
Powerful	172.38	370.019	.562	.884
Creative	171.61	374.878	.411	.887
Trustworthy	169.79	389.528	.373	.888
Can concentrate well on a task	170.74	372.832	.503	.885
Warm	170.72	379.260	.403	.887
Tough	171.96	374.438	.422	.886
Has innate ability	171.48	375.329	.498	.885
Easily embarrassed (reverse scored)	172.35	371.548	.412	.887
Lacks confidence (reverse scored)	172.25	362.717	.591	.883
Easily hurt (reverse scored)	172.27	375.826	.379	.887
Self-conscious (reverse scored)	173.30	370.917	.491	.885
Makes mistakes (reverse scored)	172.33	372.860	.486	.885
Easily rattled (reverse scored)	172.47	370.933	.494	.885

Reliability statistics for SFSCS subscale Likeability: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.851	.858	6

Item-Total statistics for SFSCS subscale Likeability: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Fun to be with	27.42	22.401	.524	.847
Friendly	26.66	21.523	.743	.811
Sociable	27.87	18.664	.657	.828
Easy to talk to	26.79	21.331	.605	.832
Pleasant	26.60	22.836	.612	.833
Warm	26.88	20.019	.740	.806

Reliability statistics for SFSCS subscale Task Accomplishment: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.829	.835	6

Item-Total statistics for SFSCS subscale Task Accomplishment: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Hard worker	27.65	21.988	.535	.815
Productive	27.99	20.670	.620	.798
Plans ahead	28.03	20.318	.479	.832
Works efficiently	27.87	19.411	.783	.767
Good at meeting deadlines	27.61	20.021	.599	.802
Can concentrate well on a task	27.86	19.288	.633	.795

Reliability statistics for SFSCS subscale Power: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.820	.819	7

Item-Total statistics for SFSCS subscale Power: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Dominant	23.58	29.082	.597	.790
Strong	22.02	32.263	.448	.814
Forceful	24.36	29.793	.559	.797
Acts as a leader	23.01	28.516	.584	.792
Aggressive	24.65	31.086	.443	.816
Powerful	23.45	27.371	.725	.768
Tough	23.02	28.109	.579	.794

Reliability statistics for SFSCS subscale Vulnerability: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.859	.859	6

Item-Total statistics for SFSCS subscale Vulnerability: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Easily embarrassed	21.21	31.133	.668	.832
Lacks confidence	21.30	31.972	.654	.834
Easily hurt	21.28	32.491	.654	.834
Self-conscious	20.25	32.585	.688	.828
Makes mistakes when flustered	21.23	35.365	.540	.853
Easily rattled when people are watching	21.09	32.608	.692	.827

Reliability statistics for SFSCS subscale Gifted: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.833	.833	5

Item-Total statistics for SFSCS subscale Gifted: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
A natural talent	18.55	16.360	.691	.782
Has special talents	18.71	15.440	.707	.777
Bright and ingenious	17.67	18.991	.543	.823
Creative	18.05	16.206	.618	.805
Has innate ability	17.92	17.785	.615	.804

Reliability statistics for SFSCS subscale Moral: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.712	.732	6

Item-Total statistics for SFSCS subscale Moral: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Loyal	31.29	6.407	.425	.680
Law-abiding	31.49	6.692	.330	.714
Truthful	31.52	7.043	.448	.674
Faithful	31.33	6.442	.390	.694
Honest	31.38	6.502	.625	.627
Trustworthy	31.14	6.694	.542	.648

Appendix 24 - Component structure for SFSCS – Sub-project B

KMO and Bartlett's Test: Before completion of personality assessment

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.774
Bartlett's Test of Sphericity	Approx. χ^2	1733.770
	df	630
	Sig.	.000

Component structure and proportion of variance explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	7.970	22.139	22.139	7.970	22.139	22.139	3.890
2	4.230	11.750	33.889	4.230	11.750	33.889	2.205
3	3.006	8.350	42.240	3.006	8.350	42.240	4.330
4	2.424	6.734	48.974	2.424	6.734	48.974	4.887
5	2.045	5.682	54.656	2.045	5.682	54.656	2.366
6	1.901	5.280	59.935	1.901	5.280	59.935	4.379
7	1.223	3.396	63.332	1.223	3.396	63.332	2.914
8	1.101	3.058	66.390	1.101	3.058	66.390	4.431
9	1.022	2.838	69.228	1.022	2.838	69.228	1.187
Extraction Method: Principal Component Analysis.							
a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.							

Pattern Matrix

	Component								
	1	2	3	4	5	6	7	8	9
Fun to be with	.129	-.068	-.676	-.046	-.180	-.034	-.001	.126	.140
Hard worker	.176	.248	-.201	.008	-.165	-.146	-.232	-.441	.350
Dominant	.543	-.191	.016	-.334	-.001	-.036	-.058	.105	.146
Easily embarrassed	-.019	.068	-.047	.828	.026	-.105	-.112	.088	.251
A natural talent	.089	.116	.012	-.009	.023	-.861	-.011	.095	.268
Loyal	.192	.092	-.026	-.104	.320	.133	-.598	.065	.145
Strong	.572	.104	.036	-.071	-.110	-.124	-.191	-.181	-.021
Friendly	-.112	.016	-.810	.042	-.006	.031	-.122	-.116	.110
Productive	.312	.123	-.116	-.164	.039	-.083	.186	-.623	.137
Lacks confidence	-.055	.118	.211	.674	.174	.127	-.060	.084	-.032
Law-abiding	-.110	.247	-.061	.139	.478	-.174	.120	-.278	-.265
Forceful	.257	-.729	.093	-.168	.010	-.042	-.098	.000	.015
Has special talents	-.013	-.020	.056	-.091	.034	-.800	.058	-.096	.005
Plans ahead	-.426	-.227	-.059	-.033	.063	-.214	-.046	-.678	.172
Sociable	.017	-.208	-.850	-.027	.132	-.023	.210	.103	-.044
Easily hurt	-.045	-.281	-.175	.755	.049	.096	-.033	.045	-.158
Acts as a leader	.454	-.062	-.166	-.319	-.084	-.217	-.109	.012	-.022
Truthful	.036	-.095	-.034	-.058	.878	-.007	-.009	.045	.180
Self-conscious	.161	.187	-.007	.753	-.059	.224	.054	-.071	.064
Works efficiently	.155	-.026	.040	-.023	-.009	-.021	-.094	-.814	.039
Faithful	-.152	-.216	-.073	.152	.009	-.244	-.676	-.267	.071
Aggressive	.476	-.570	.026	.136	.006	.066	.294	-.049	-.027

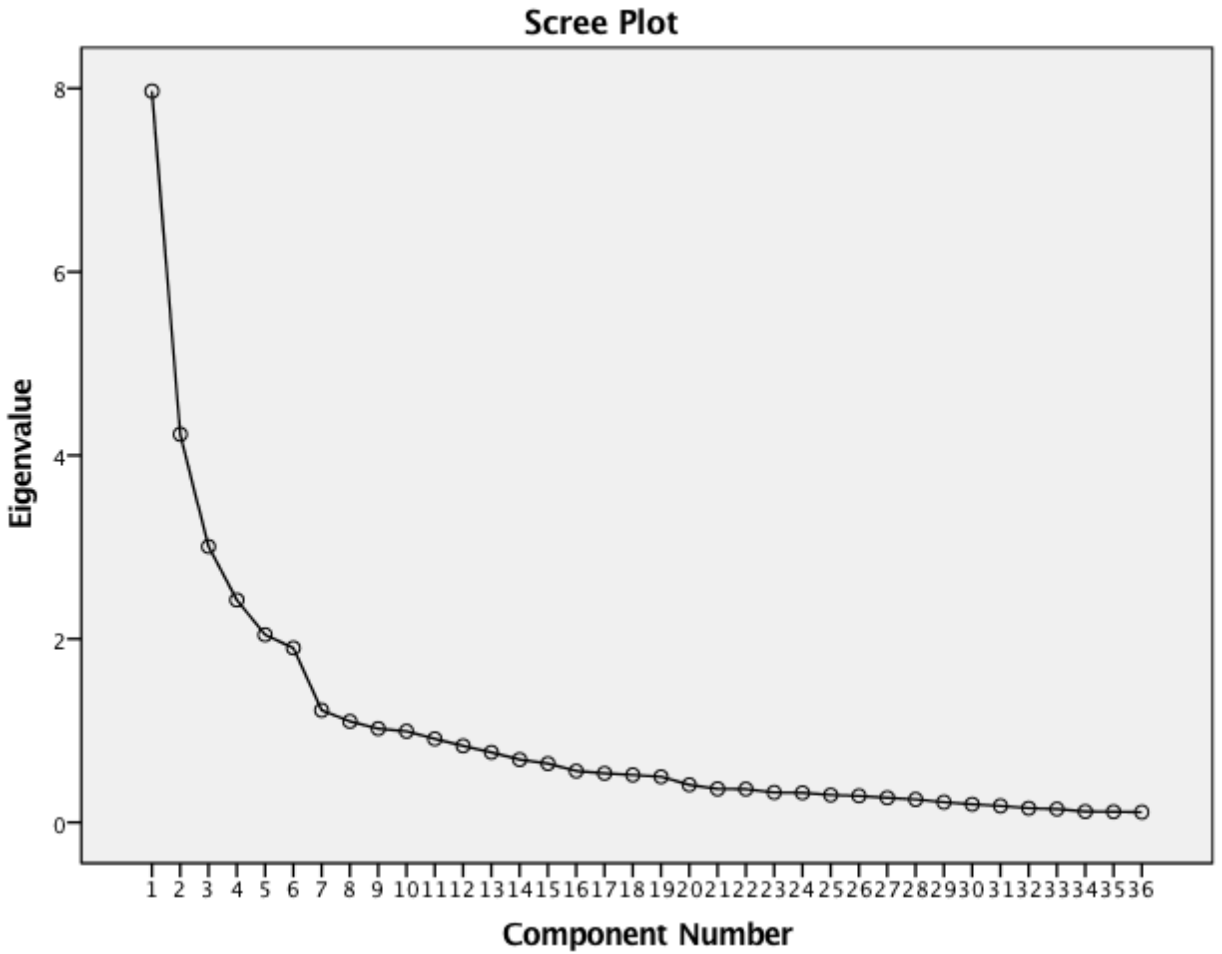
Easy to talk to	.008	-.047	-.634	-.209	.030	.065	-.281	.086	-.278
Bright and ingenious	.152	.114	-.025	-.168	-.109	-.462	-.205	-.040	-.479
Makes mistakes when flustered	-.080	-.176	.055	.621	-.243	-.030	-.068	.211	-.189
Honest	-.019	-.046	.013	-.005	.802	-.038	-.233	.047	-.147
Good at meeting deadlines	-.070	-.006	.010	.020	.052	.116	-.096	-.796	-.085
Pleasant	.033	.339	-.693	.091	.100	.015	.026	-.137	-.118
Powerful	.623	-.282	-.180	-.068	.004	-.148	.116	-.138	.035
Creative	-.119	-.232	-.138	.033	-.062	-.781	.003	.028	-.135
Easily rattled when people are watching	.016	.049	.125	.803	-.035	-.058	.169	-.092	-.011
Trustworthy	.091	.057	-.129	.014	.151	.087	-.642	-.167	-.223
Can concentrate well on a task	.264	.072	.053	-.117	-.119	.019	-.075	-.672	-.320
Warm	-.036	.220	-.786	.003	.036	-.011	-.119	-.088	.012
Tough	.621	-.280	-.031	.134	.049	-.144	-.084	-.101	-.075
Has innate ability	.275	.168	.104	-.034	.159	-.678	.000	-.008	-.193
Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalisation.									

Structure Matrix

	Component								
	1	2	3	4	5	6	7	8	9
Fun to be with	.199	-.130	-.663	-.166	-.167	-.108	-.092	.025	.154
Hard worker	.259	.270	-.331	-.161	-.078	-.304	-.362	-.586	.323
Dominant	.651	-.289	-.066	-.495	-.066	-.194	-.083	.016	.146
Easily embarrassed	-.241	.086	.030	.794	.033	.097	-.018	.098	.213
A natural talent	.214	.021	-.082	-.215	.014	-.834	-.089	-.154	.232
Loyal	.189	.140	-.178	-.192	.399	.031	-.650	-.100	.095
Strong	.641	.044	-.088	-.280	-.088	-.300	-.268	-.333	-.046
Friendly	-.050	.052	-.836	-.045	.085	-.068	-.296	-.235	.103
Productive	.429	.140	-.224	-.311	.061	-.316	-.020	-.696	.122
Lacks confidence	-.310	.170	.299	.748	.176	.327	.054	.167	-.062
Law-abiding	-.140	.327	-.104	.132	.540	-.187	-.061	-.370	-.308
Forceful	.406	-.763	.021	-.271	-.066	-.185	-.061	.033	.013
Has special talents	.163	-.075	-.048	-.254	.039	-.831	-.052	-.291	-.026
Plans ahead	-.258	-.089	-.174	-.046	.140	-.344	-.181	-.668	.149
Sociable	.085	-.238	-.808	-.112	.131	-.098	.036	.020	-.035
Easily hurt	-.220	-.241	-.076	.763	.049	.225	.040	.135	-.180
Acts as a leader	.612	-.154	-.278	-.520	-.087	-.401	-.211	-.163	-.030
Truthful	.001	-.027	-.104	-.079	.855	-.032	-.143	-.042	.125
Self-conscious	-.099	.208	.103	.761	-.059	.361	.135	.021	.047
Works efficiently	.280	.067	-.120	-.145	.069	-.274	-.260	-.850	.008
Faithful	-.038	-.122	-.257	.026	.141	-.353	-.728	-.416	.013
Aggressive	.491	-.639	.064	.035	-.123	-.017	.324	.055	-.019

Easy to talk to	.115	-.041	-.701	-.289	.126	-.077	-.432	-.073	-.283
Bright and ingenious	.304	.065	-.149	-.317	-.031	-.571	-.317	-.260	-.502
Makes mistakes when flustered	-.227	-.190	.166	.653	-.254	.153	.099	.302	-.190
Honest	-.047	.041	-.091	-.023	.842	-.069	-.362	-.084	-.213
Good at meeting deadlines	.014	.134	-.114	-.009	.154	-.090	-.250	-.779	-.112
Pleasant	.024	.359	-.702	.008	.196	-.061	-.189	-.288	-.131
Powerful	.730	-.372	-.250	-.298	-.056	-.342	.011	-.230	.024
Creative	.060	-.290	-.206	-.120	-.049	-.784	-.077	-.144	-.158
Easily rattled when people are watching	-.201	.066	.239	.814	-.056	.124	.259	.008	-.032
Trustworthy	.130	.139	-.296	-.082	.295	-.064	-.740	-.335	-.275
Can concentrate well on a task	.387	.123	-.086	-.228	-.034	-.238	-.232	-.726	-.338
Warm	.011	.242	-.824	-.097	.143	-.110	-.325	-.262	-.001
Tough	.668	-.346	-.116	-.092	.015	-.295	-.141	-.201	-.107
Has innate ability	.374	.093	-.008	-.230	.171	-.722	-.127	-.254	-.237
Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalisation.									

Scree plot


Scree Plot
Eigenvalue
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
Component Number

Appendix 25 - Correlations between self-concept variables (SFSCS subscale scores) and self-esteem (RSES total score) – Sub-project B

Before completion of personality assessment

Sub-project B - before completion of personality assessment		Likeability	Task Accomplishment	Power (weighted)	Vulnerability	Gifted (weighted)	Moral	Self-esteem
SELF-CONCEPT Likeability	Pearson's r	1	.248*	.15	-.199	.149	.319**	.250*
	Sig.(2-tailed)		.017	.153	.057	.157	.002	.016
	N	92	92	92	92	92	92	92
Task Accomplishment	Pearson's r	.248*	1	.262*	-.263*	.385**	.389**	.462**
	Sig.(2-tailed)	.017		.012	.011	0	0	0
	N	92	92	92	92	92	92	92
Power (weighted)	Pearson's r	.15	.262*	1	-.416**	.410**	.031	.454**
	Sig.(2-tailed)	.153	.012		0	0	.77	0
	N	92	92	92	92	92	92	92
Vulnerability	Pearson's r	-.199	-.263*	-.416**	1	-.344**	-.09	-.573**
	Sig.(2-tailed)	.057	.011	0		.001	.393	0
	N	92	92	92	92	92	92	92
Gifted (weighted)	Pearson's r	.149	.385**	.410**	-.344**	1	.220*	.571**
	Sig.(2-tailed)	.157	0	0	.001		.035	0
	N	92	92	92	92	92	92	92
Moral	Pearson's r	.319**	.389**	.031	-.09	.220*	1	.263*
	Sig.(2-tailed)	.002	0	.77	.393	.035		.011
	N	92	92	92	92	92	92	92
SELF-ESTEEM Self-esteem	Pearson's r	.250*	.462**	.454**	-.573**	.571**	.263*	1
	Sig.(2-tailed)	.016	0	0	0	0	.011	
	N	92	92	92	92	92	92	92

After completion of personality assessment

Sub-project B - after completion of personality assessment		Likeability	Task Accomplishment	Power (weighted)	Vulnerability	Gifted (weighted)	Moral	Self-esteem
SELF-CONCEPT Likeability	Pearson's r	1	.342**	.209*	-.198	.295**	.436**	.334**
	Sig. (2-tailed)		.001	.045	.059	.004	0	.001
	N	92	92	92	92	92	92	92
Task Accomplishment	Pearson's r	.342**	1	.129	-.277**	.440**	.530**	.448**
	Sig. (2-tailed)	.001		.222	.007	0	0	0
	N	92	92	92	92	92	92	92
Power (weighted)	Pearson's r	.209*	.129	1	-.299**	.316**	.056	.343**
	Sig. (2-tailed)	.045	.222		.004	.002	.593	.001
	N	92	92	92	92	92	92	92
Vulnerability	Pearson's r	-.198	-.277**	-.299**	1	-.204	-.243*	-.537**
	Sig. (2-tailed)	.059	.007	.004		.051	.02	0
	N	92	92	92	92	92	92	92
Gifted (weighted)	Pearson's r	.295**	.440**	.316**	-.204	1	.293**	.448**
	Sig. (2-tailed)	.004	0	.002	.051		.005	0
	N	92	92	92	92	92	92	92
Moral	Pearson's r	.436**	.530**	.056	-.243*	.293**	1	.314**
	Sig. (2-tailed)	0	0	.593	.02	.005		.002
	N	92	92	92	92	92	92	92
SELF-ESTEEM Self-esteem	Pearson's r	.334**	.448**	.343**	-.537**	.448**	.314**	1
	Sig. (2-tailed)	.001	0	.001	0	0	.002	
	N	92	92	92	92	92	92	92

Appendix 26 - Descriptive statistics for self-esteem (RSES total score) – Sub-project B

Before and after completion of personality assessment – Sub-project B

	N	Range	Minimum	Maximum	x̄		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Before completion of personality assessment	92	30	0	30	20.60	.613	5.876	34.529	-.345	.251	.161	.498
After completion of personality assessment	92	25	5	30	21.13	.585	5.609	31.455	-.167	.251	-.659	.498
Valid N	92											

Tests of Normality: Before and after completion of personality assessment – Sub-project B

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Before completion of personality assessment	.093	92	.049	.956	92	.004
After completion of personality assessment	.125	92	.001	.960	92	.006
a. Lilliefors Significance Correction						

Appendix 27 – Reliability analysis for RSES – Sub-project B

Reliability statistics for RSES: Before completion of personality assessment – Sub-project B

Cronbach's α	Cronbach's α based on standardised items	N of items
.921	.924	10

Item-Total statistics for RSES: Before completion of personality assessment – Sub-project B

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's α if Item Deleted
I feel that I am a person of worth, at last on an equal plane with others	18.21	30.100	.628	.545	.917
I feel that I have a number of good qualities	18.12	29.227	.787	.675	.911
All in all, I am inclined to feel that I am a failure	18.39	27.647	.784	.652	.908
I am able to do things as well as most other people	18.34	30.314	.554	.479	.920
I feel I do not have much to be proud of	18.39	28.087	.724	.658	.912
I take a positive attitude toward myself	18.59	27.806	.781	.666	.909
On the whole, I am satisfied with myself	18.55	27.503	.749	.655	.910
I wish I could have more respect for myself	18.91	26.322	.740	.586	.912
I certainly feel useless at times	19.16	27.720	.687	.583	.914
At times I think I am no good at all	18.72	27.414	.669	.619	.916

Appendix 28 – Component structure for RSES – Sub-project B

KMO and Bartlett's Test: Before completion of personality assessment

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.905
Bartlett's Test of Sphericity	Approx. χ^2	585.945
	df	45
	Sig.	.000

Component structure and proportion of variance explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	5.989	59.889	59.889	5.989	59.889	59.889	4.945
2	1.050	10.501	70.390	1.050	10.501	70.390	4.766
Extraction Method: Principal Component Analysis.							
a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.							

Appendix 29 – Descriptive statistics for self-concept and self-esteem change scores – Sub-project B

	N	Range	Minimum	Maximum	x̄		SD	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SELF-CONCEPT Change in Likeability	92	15	-6	9	.61	.304	2.912	8.483	.267	.251	.236	.498
Change in Task Accomplishment	92	17	-8	9	.03	.292	2.799	7.834	.190	.251	.915	.498
Change in Power	92	24.00	-9.43	14.57	.009	.391	3.753	14.088	.163	.251	1.696	.498
Change in Vulnerability	92	21	-10	11	-.74	.401	3.848	14.810	-.102	.251	.402	.498
Change in Gifted	92	16.80	-8.40	8.40	.561	.379	3.635	13.212	-.147	.251	-.328	.498
Change in Moral	92	16	-8	8	-.32	.249	2.390	5.713	-.160	.251	1.999	.498
Change in Total self-concept	92	55.94	-28.43	27.51	1.635	1.115	10.696	114.410	-.023	.251	.039	.498
SELF-ESTEEM Change in Self-esteem	92	16	-10	6	.53	.268	2.570	6.603	-.736	.251	2.366	.498
Valid N	92											

Appendix 30 - Normality of distributions for self-concept and self-esteem change scores - Sub-project B

Kolmogorov-Smirnov test

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Change in Likeability	.120	92	.002
Change in Task Accomplishment	.093	92	.048
Change in Power (weighted)	.111	92	.007
Change in Vulnerability	.114	92	.005
Change in Gifted (weighted)	.134	92	.000
Change in Moral	.150	92	.000
Change in Total self-concept (weighted)	.050	92	.200[*]
Change in Self-esteem	.124	92	.001
*. This is a lower bound of the true significance.			
a. Lilliefors Significance Correction			

Appendix 31 – Paired samples t-tests for self-concept and self-esteem scores – Sub-project B

	Paired Differences					t	df	Sig. (2-tailed)
			Std. Error	95% Confidence Interval of the Difference				
	x̄	SD	Mean	Lower	Upper			
Likeability 1 versus Likeability 2	-.609	2.912	.304	-1.212	-.006	-2.005	91	.048
Task Accomplishment 1 versus Task Accomplishment 2	-.033	2.799	.292	-.612	.547	-.112	91	.911
Power 1 (weighted) versus Power 2 (weighted)	-.009	3.753	.391	-.787	.768	-.024	91	.981
Vulnerability 1 versus Vulnerability 2	.739	3.848	.401	-.058	1.536	1.842	91	.069
Gifted 1 (weighted) versus Gifted 2 (weighted)	-.561	3.635	.379	-1.314	.192	-1.480	91	.142
Moral 1 versus Moral 2	.315	2.390	.249	-.180	.810	1.265	91	.209
Total Self-concept 1 (weighted) versus Total Self-concept 2 (weighted)	-1.635	10.696	1.115	-3.851	.580	-1.467	91	.146
Self-esteem 1 versus Self-esteem 2	-.533	2.570	.268	-1.065	.000	-1.988	91	.050

Appendix 32 – Correlations between interval-level variables and change scores – Sub-project B

		Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.
SELF-CONCEPT Change in Likeability	Pearson's r	.083	.001
	Sig. (2-tailed)	.432	.992
	N	92	92
Change in Task Accomplishment	Pearson's r	.209*	-.093
	Sig. (2-tailed)	.045	.380
	N	92	92
Change in Power (weighted)	Pearson's r	.091	.009
	Sig. (2-tailed)	.390	.931
	N	92	92
Change in Vulnerability	Pearson's r	.111	.043
	Sig. (2-tailed)	.292	.682
	N	92	92
Change in Gifted (weighted)	Pearson's r	.141	-.070
	Sig. (2-tailed)	.181	.510
	N	92	92
Change in Moral	Pearson's r	-.054	.194
	Sig. (2-tailed)	.611	.064
	N	92	92
Change in Total Self-concept (weighted)	Pearson's r	.105	-.017
	Sig. (2-tailed)	.319	.876
	N	92	92

		Overall, do you expect completing the personality assessment to be a positive or negative experience for you?	Please select below from 1 to 5 according to your confidence that your personality assessment results will be correctly interpreted.
SELF-ESTEEM Change in Self-esteem	Pearson's r	.130	-.045
	Sig. (2-tailed)	.216	.673
	N	92	92

Appendix 33 –Analysis of variance for potential predictors of self-perception change scores (MANOVA) – Sub-project B

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.071	.756[a]	7.000	69.000	.626
	Wilks' Lambda	.929	.756[a]	7.000	69.000	.626
	Hotelling's Trace	.077	.756[a]	7.000	69.000	.626
	Roy's Largest Root	.077	.756[a]	7.000	69.000	.626
Are you female or male?	Pillai's Trace	.053	.556[a]	7.000	69.000	.789
	Wilks' Lambda	.947	.556[a]	7.000	69.000	.789
	Hotelling's Trace	.056	.556[a]	7.000	69.000	.789
	Roy's Largest Root	.056	.556[a]	7.000	69.000	.789
What is your ethnicity?	Pillai's Trace	.149	.804	14.000	140.000	.664
	Wilks' Lambda	.856	.798[a]	14.000	138.000	.670
	Hotelling's Trace	.163	.792	14.000	136.000	.676
	Roy's Largest Root	.117	1.167[b]	7.000	70.000	.333
What is your age?	Pillai's Trace	.482	1.113	35.000	365.000	.307
	Wilks' Lambda	.595	1.098	35.000	292.687	.330
	Hotelling's Trace	.560	1.079	35.000	337.000	.354
	Roy's Largest Root	.232	2.422[b]	7.000	73.000	.027
Are you female or male? * What is your ethnicity?	Pillai's Trace	.036	.372[a]	7.000	69.000	.916
	Wilks' Lambda	.964	.372[a]	7.000	69.000	.916
	Hotelling's Trace	.038	.372[a]	7.000	69.000	.916
	Roy's Largest Root	.038	.372[a]	7.000	69.000	.916
Are you female or male?	Pillai's Trace	.391	1.522	21.000	213.000	.072

* What is your age?	Wilks' Lambda	.653	1.516	21.000	198.681	.075
	Hotelling's Trace	.467	1.504	21.000	203.000	.079
	Roy's Largest Root	.246	2.499[b]	7.000	71.000	.024
What is your ethnicity? * What is your age?	Pillai's Trace	.368	1.419	21.000	213.000	.111
	Wilks' Lambda	.672	1.402	21.000	198.681	.120
	Hotelling's Trace	.429	1.383	21.000	203.000	.130
	Roy's Largest Root	.227	2.300[b]	7.000	71.000	.036
Are you female or male? * What is your ethnicity? * What is your age?	Pillai's Trace	.000	.[a]	.000	.000	.
	Wilks' Lambda	1.000	.[a]	.000	72.000	.
	Hotelling's Trace	.000	.[a]	.000	2.000	.
	Roy's Largest Root	.000	.000[a]	7.000	68.000	1.000
a. Exact statistic						
b. The statistic is an upper bound on F that yields a lower bound on the significance level.						

Appendix 34 - Correlations between self-concept and self-esteem change scores - Sub-project B

		Change in Likeability	Change in Task Accomplishment	Change in Power	Change in Vulnerability	Change in Gifted	Change in Moral	Change in Self-esteem
SELF-CONCEPT Change in Likeability	Pearson's r	1	.402**	.179	-.153	.329**	.249*	.182
	Sig. (2-tailed)		.000	.088	.147	.001	.017	.082
	N	92	92	92	92	92	92	92
Change in Task Accomplishment	Pearson's r	.402**	1	.253*	-.092	.374**	.328**	.175
	Sig. (2-tailed)	.000		.015	.385	.000	.001	.096
	N	92	92	92	92	92	92	92
Change in Power (weighted)	Pearson's r	.179	.253*	1	.119	.316**	-.114	.144
	Sig. (2-tailed)	.088	.015		.258	.002	.279	.171
	N	92	92	92	92	92	92	92
Change in Vulnerability	Pearson's r	-.153	-.092	.119	1	-.044	-.028	-.239*
	Sig. (2-tailed)	.147	.385	.258		.680	.791	.022
	N	92	92	92	92	92	92	92
Change in Gifted (weighted)	Pearson's r	.329**	.374**	.316**	-.044	1	.104	.250*
	Sig. (2-tailed)	.001	.000	.002	.680		.324	.016
	N	92	92	92	92	92	92	92
Change in Moral	Pearson's r	.249*	.328**	-.114	-.028	.104	1	.010
	Sig. (2-tailed)	.017	.001	.279	.791	.324		.927
	N	92	92	92	92	92	92	92
SELF-ESTEEM Change in Self-esteem	Pearson's r	.182	.175	.144	-.239*	.250*	.010	1
	Sig. (2-tailed)	.082	.096	.171	.022	.016	.927	
	N	92	92	92	92	92	92	92