An Evaluation of the Personality Assessment Inventory for Personnel Selection: A Study with the New Zealand Police

A thesis presented in partial fulfilment of
the requirements for the degree of
Master of Arts in Psychology
at Massey University, Palmerston North,
New Zealand.

Nick Lascelles

2000

Abstract

The Personality Assessment Inventory (PAI) was evaluated as an instrument for New Zealand police selection. Recruits (N = 127) attending training were administered the PAI, and the data used to assess reliability, validity, impact of response sets, test fairness, and utility for selection purposes. The PAI was found to be less internally consistent than in other studies, but most scales achieved acceptable alphas. The PAI showed evidence of construct validity when correlated with the NEO-PI-R. PAI scores accounted for 12% of the variance in final grade achieved during training, compared to 8% for the NEO-PI-R. Defensive responding was identified as a considerable threat to validity. Small, but significant differences were found between demographic groups on some scales based on age and ethnicity. While the PAI might be a useful addition to the New Zealand Police test battery, the presence of less than ideal reliability, possibly high levels of defensive responding, and differences between demographic groups means that hypotheses generated from test scores will need to be carefully integrated with other sources of information.

Acknowledgements

I would like to thank:-

Dr Ross Flett for his patience and assistance as my supervisor.

The members of Wings 186 and 187 of the Royal New Zealand Police College for taking part in this research.

Jonathan Black for his help in facilitating the involvement of the New Zealand Police.

Sergeants' Sarah Stirling and Steve Le Fevre for providing time in the Royal New Zealand Police College's busy schedule to test recruits.

Manaia Lascelles for motivating me to finish this project.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	viii
List of Figures	xii
Chapter 1: The Background of This Study	1
1.0: Introduction	1
1.1: The Current Study	3
1.2: Overview	3
Chapter 2: Psychological Consequences of Policing	4
Chapter 3: Police Selection	7
3.0: History	7
3.1: Criticism of Personality Test Use in Police Selection	7
3.2: Screen in/Screen out	10
3.3: New Zealand Police Selection	13
Chapter 4: A Review of Tests Used in This Study	17
4.0: PAI	17
4.1: NEO-PI-R	23
Chapter 5: Special Issues in Test Evaluation	24
5.0: Overview	24
5.1: Criterion-Related Validity	24
5.2: Response Distortion	26
5.3: Test Fairness	32

Chapter 6: The Current Study 36

Chapter 7: Methodology 38
7.0: Participants 38
7.0.0: Sample Selection 39
7.1: Information Sheet 40
7.2: Test Administration 40
7.3: Ethical Issues and Confidentiality 41
7.4: Data Analysis 42
7.5: Test Interpretation 43
7.6: Measures 44
7.6.0: PAI 44
7.6.1: NEO-PI-R 48
7.6.2: RNZPC Performance Data 49

Chapter 8: Results 50
8.0: Overview 50
8.1: Reliability 51
8.2: Validity of PAI Test Protocols 55
8.3: Construct Validity 57
8.3.0: Interscale Correlation 57
8.3.1: Structure of the PAI 60
8.4: Criterion-Related Predictive Validity 63
8.4.0: Effect of Response Sets on Predictive Validity 63
8.4.1: Prediction of RNZPC Final Grade with the PAI and NEO-PI-R 66
8.5: Test Fairness 76
8.5.0: Univariate Correlations Between Demographics and Performance 76
8.5.1: The Four-Fifths Rule 79
8.6: PAI Scores for the Police Recruit Sample 80
8.7: Potential Danger to Self and Others 86
8.7.0: Potential for Aggression 86

8.7.1: Potential for Self-Harm 87
8.8: Elevations in Clinical and Treatment Scales 88

Chapter 9: Discussion 89
9.0: Overview 89
9.1: Psychometric Properties of the PAI 90
9.1.0: Reliability 90
9.1.1: Protocol Validity 91
9.1.2: Effects of Response Sets on Predictive Validity 94
9.1.3: Construct Validity 96
9.1.4: Structure of the PAI 96
9.1.5: Criterion-Related Predictive Validity 97
9.1.6: Prediction of RNZPC Final Grade with the PAI and NEO-PI-R 98
9.2: Test Fairness 100
9.2.0: Age 100
9.2.1: Gender 101
9.2.2: Ethnicity 102
9.2.3: The Four-Fifths Rule 103
9.3: PAI Scores for the Police Recruit Sample 105
9.4: Potential Danger to Self and Others 107
9.5: Elevations in PAI Clinical and Treatment Scales 108
9.6: NEO-PI-R vs PAI 111
9.7: Limitations 112
9.8: Directions for Future Research 114
9.9: Conclusions 115

References 117

Appendices 129
A: Data 129
B: Information Sheet 138
C: Test Scale Descriptions 141

List of Tables

Table 1. Demographic Details of Police Recruits in the Sample (*N* = 127). 38

Table 2. Demographic Details of all Police Recruits Asked to Participate in the Study (*N* = 160). 39

Table 3. Criteria for a Valid PAI Protocol in the Present Study. 46

Table 4. Alpha Coefficients and Mean Inter-item Correlation for PAI Full Scales for the Police Recruit (*N* = 127) and Community Samples (*N* = 1000). 53

Table 5. Alpha Coefficients and Mean Inter-item Correlation for PAI Subscales for the Police Recruit (*N* = 127) and Community Samples (*N* = 1000). 54

Table 6. Validity of PAI Protocols for the Police Recruit Sample (*N* = 127). 55

Table 7. Significant Differences Between Mean PAI Scale Scores for Valid and Invalid Protocols for the Police Recruit Sample (*N* = 127). 56

Table 8. Orthogonal PAI Factor Pattern for the Police Recruit Sample (*N* = 124). 62

Table 9. Correlations Between Final Grade Obtained at the RNZPC and Scores on the PAI and NEO-PI-R for all Recruits and Valid and Invalid Protocols. 64

Table 10. Hierarchical Multiple Regression of Demographic, Protocol 67

Validity and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 124).

Table 11. Hierarchical Multiple Regression of Demographic, Protocol Validity and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 121). 68

Table 12. Hierarchical Multiple Regression of Demographic, Protocol Validity and NEO-PI-R Domain Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} (N = 106). 69

Table 13. Hierarchical Multiple Regression of Demographic, Protocol Validity and NEO-PI-R Facet Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} (N = 106). 70

Table 14. Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Domain Scales, and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 102). 71

Table 15. Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Facet Scale, and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for 72

Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 101).

Table 16. Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Facet Scale, and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 102). 73

Table 17. Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Domain Scales, and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 101). 74

Table 18. Significant Differences on PAI Scales Between Male (N = 104) and Female (N = 23) Police Recruits. 77

Table 19. Significant Differences on NEO-PI-R Scales Between Male (N = 104) and Female (N = 23) Police Recruits. 77

Table 20. Significant Differences in Test Scores Between Maori (N = 17) and Caucasian (N = 110) Police Recruits. 78

Table 21. Significant Differences in Test Scores Between Partnered (N = 62) and Unpartnered Police Recruits (N = 65). 78

Table 22. PAI Full Scale T scores for the Police Recruit Sample (N = 127). 80

Table 23. PAI Subscale *T* Scores for the Police Recruit Sample (*N* = 127). 82

Table 24. Elevations on PAI Scales and Indexes Related to Potential for Aggression for the Police Recruit Sample (*N* =127). 86

Table 25. PAI Full Scale Raw Scores for the Police Recruit Sample (*N* = 127). 129

Table 26. PAI Subscale Raw Scores for the Police Recruit Sample (*N* = 127). 130

Table 27. Intercorrelation Matrix for PAI Full Scales for the Police Recruit Sample (*N* = 127). 131

Table 28. Intercorrelation Between PAI Full and Subscales for the Police Recruit Sample (*N* = 127). 132

Table 29. Intercorrelation Between PAI Full Scales and NEO-PI-R Facet Scales for the Police Recruit Sample (*N* = 106). 133

Table 30a. Intercorrelation Between PAI Subscales and NEO-PI-R Facet Scales for the Police Recruit Sample (*N* = 106). 134

Table 30b. Intercorrelation Between PAI Subscales and NEO-PI-R Facet Scales for the Police Recruit Sample (*N* = 106). 135

Table 31. Intercorrelation Between PAI Subscales and NEO-PI-R Domain Scales for the Police Recruit Sample (*N* = 106). 136

Table 32. Intercorrelation Between PAI Full Scales and NEO-PI-R Domain Scales (*N* = 106) 137

Table 33. Description of PAI Full Scales (Morey, 1996). 141

Table 34a. Description of PAI Subscales (Morey, 1996). 142

Table 34b. Description of PAI Subscales (Morey, 1996). 143

Table 35a. Description of NEO-PI-R Domain and Facet Scales (Costa & McCrae, 1992a). 144

Table 35b. Description of NEO-PI-R Domain and Facet Scales (Costa & McCrae, 1992a). 145

List of Figures

Figure 1. Comparison of Mean PAI Full Scale Recruit Profile (*N* = 127) With Normative Community Sample (*N* = 1000). 84

Figure 2. Comparison of Mean PAI Subscale Recruit Profile (*N* = 127) With Normative Community Sample (*N* =1000). 85

Chapter 1: The Background of this Study

1.0 Introduction

The selection of suitable candidates for training as police officers is a vital function of a law enforcement agency. The duties and responsibilities of a police officer mean that unsuitable officers can cause considerable damage to an agency's relationship with the public, and even place the lives of civilians and fellow staff at risk (Leonard & More, 1993).

The role of a police officer requires the exercise of considerable discretionary power in interpreting and applying the law, including the ability to detain or arrest persons. A police officer must have the ability to cope with a variety of situations that may involve physical danger and are emotionally traumatic, and requires a high degree of interpersonal skills in interacting with the public. As well as dealing with more dramatic situations, the officer must cope with working within the structure of a paramilitary organisation, dealing with administrative duties, and being constantly in the public eye (Beutler, Nussbaum & Meredith, 1988).

The New Zealand Police comprised 6492 sworn staff and 1909 nonsworn staff in 1997 (New Zealand Police, 1998). With attrition rates of about six percent in recent years, and sworn staff numbers increasing as a result of government initiatives, a substantial number of new staff are required to be recruited and trained annually.

The training of an individual officer is a costly investment running into many thousands of dollars and a considerable waste if an unsuitable officer exits from employment prematurely, either by termination or voluntarily. The total cost of training and employing a general duties constable through to the end of his/her two year probationary period is estimated to be $144,000 (New Zealand Police, 1998). There are then considerable savings to be made if

psychologically unsuitable applicants can be identified early in the selection process.

The use of interviews, work experience placements, and background investigations may screen out some individuals with undesirable psychological characteristics. However, Leonard and More (1993) state that psychological screening is a necessity in order to detect latent personality problems such as personality disorders that are not easily identified.

The use of psychological tests to select police personnel has become commonplace (Fyfe, Greene, Walsh, Wilson, & McLaren, 1997). The type of test used varies, with most batteries including a test of cognitive ability, and some measure of personality that is intended to measure desirable or undesirable psychological characteristics.

In the past, personality tests have been employed primarily to identify applicants whose personality defects may make them a liability to the organisation. In recent years, there has been increased focus on assessing normal personality characteristics in order to select the best personnel, as opposed to screening out the worst. The work of Barrick and Mount (1991) was instrumental in demonstrating solid relationships between personality characteristics such as conscientiousness and job performance across a number of occupations including police. Cortina, Doherty, Schmitt, Kaufman, and Smith (1992) have also found substantial correlations between personality characteristics and police performance.

Despite the increased emphasis on measuring normal personality characteristics, there is a call by some for the screening out of undesirable individuals to remain the primary focus of psychological assessment in the police context (Fyfe et al., 1997). They cite the relatively low predictive value of personality tests as a reason for concentrating on assessing the current psychological health of applicants.

The application of a psychological test for selection requires the test to be validated for the population for which it is to be used, and it's psychometric properties evaluated. Without this information, a test cannot be used with confidence in a particular selection context. In addition to these basic properties, Kline (1993) identifies the deliberate distorting of responses to present favourably, and the possibility of a test adversely impacting on particular demographic groups as important when considering the use of psychological testing for selection purposes.

1.1 The Current Study

The current study is being conducted with the aim of evaluating the Personality Assessment Inventory (PAI; Morey, 1991) as a measure to assist in the selection of personnel for general and special duties policing. The study includes an assessment of the test's psychometric properties, issues such as possible adverse impact on particular groups, the influence of response distortion, and an evaluation of the test's ability to screen personnel to meet the New Zealand Police standards of psychological fitness. The PAI will be administered to a group of recruits undergoing training and validated against a personality test (NEO-PI-R; Costa & McCrae, 1992a) in use currently, and performance at the Royal New Zealand Police College.

1.2 Overview

The following chapter reviews the literature related to the psychological effects of working as a police officer. Chapter 3 looks at the issues involved in police selection. Chapter 4 reviews the tests (PAI and NEO-PI-R) used in this study. Chapter 5 reviews some special issues involved in test evaluation with particular reference to aspects pertinent to the law enforcement context.

Chapter 2: Psychological Consequences of Policing

The work of police officers has been recognised as being stressful and posing some unique demands on individuals. Bartol and Bartol (1994) state that people in other occupations such as construction workers and miners may be exposed to similar or greater levels of risk of physical injury. They note, however, that while others such as paramedics may assert that they experience similar stress, few are exposed to the variety of stressors (physical, psychological, and social) found in police work. The shooting of Steven Wallace in Waitara recently highlighted how an officer's decision on a course of action can have consequences that go far beyond what most other occupations have to cope with.

Beutler et al. (1988) identifies some of the job demands that distinguish police work from other occupations. Firstly, officers are required to participate in dangerous situations on a regular basis. On an interpersonal level, the police officer is placed in contradictory roles. He/she is required on the one hand to exert a high degree of control over interactions with the public while having to comply with the instructions of superiors in a paramilitary organisation.

Not all studies have found police work to be more stressful than other occupations. Anson & Bloom (1988) found that policing was stressful, but not more so than other occupations that requires contact with criminals such as prison guards. Hart, Wearing & Headey (1995) found that police officers reported higher levels of psychological wellbeing than teachers, students and community norms. Their study showed that actual police work did little to contribute to psychological wellbeing, and it was the police organisation that was the major source of stress. Violanti & Aron (1993) support this also.

There is, however, some evidence that police work may have a detrimental effect on psychological health over time. Adlam (1982) discusses the impact of the police subculture and the nature of the work on the personality of officers

over time. He found that most officers felt that their personality had changed. In particular, there was increased self-confidence and assertiveness, emotional hardening, increased suspiciousness, and cynicism. He found that officers adhered to certain attitudes that were the norm of the subculture, as deviating from these resulted in ostracism. The norms included a distrust of ethnic minorities, dislike of the social services, and a dislike of theory and intellectuals.

Using the Minnesota Multiphasic Personality Inventory (MMPI), Beutler et al. (1988) found significant changes in a group of officers over time, with increases in somatic symptoms, anxiety and vulnerability to alcohol abuse. Robinson, Sigman, & Wilson (1997) found that 13% of a sample of 100 police officers met the DSM IV (Diagnostic and Statistical Manual of Mental Disorders; American Psychiatric Association, 1994) criteria for Post Traumatic Stress Disorder. Exposure to death and injury was identified as one of the main precursors to the development of post traumatic symptoms.

Of those staff who disengaged between 1990 and 1997 (as a result of being unable to meet requisite fitness standards) from the New Zealand Police, psychological factors were cited in 73% of cases. This amounted to 1172 of the 1656 police who disengaged during this time. Depression, stress and anxiety were most common with post traumatic reactions also reported at a high level. In some cases, more than one disorder was present (New Zealand Police, 1998).

It is interesting to note the incidence of psychological problems reported here in a population that is usually cited as having a low incidence of psychopathology (Saxe & Reiser, 1976; Beutler et al., 1988). There has been some controversy as to the veracity of some of the applications for disengagement. Claims have been made that the system encourages staff who desire a change of occupation to present with psychological complaints, in order to benefit from the provisions of the disengagement process. However, submissions from psychologists assessing police who have applied for disengagement do not

support this view. They cite many officers who continue with treatment after disengagement, and are often very reluctant to leave their career (New Zealand Police, 1998).

Although there is disagreement in the literature about whether police suffer more stress than those in other occupations, there is evidence that some police may develop psychological problems during their police career. This has some implications for selection of police personnel that are discussed in the following chapter.

Chapter 3: Police Selection

3.0 History

Psychological tests have been used to assist in selecting police for a considerable period of time. There is evidence of testing being used as early as 1917 (Poland, 1978). Narrol and Levitt (1963) reviewed the use of formal psychological testing in the United States, and found that almost all police departments were using some form of testing. However, the tests used tended to be special police selection tests that were unintentionally focused on measuring general intelligence. The personality of recruits was not assessed by most of the departments and there was little research taking place in the whole area.

In 1973 the National Advisory Commission on Criminal Justice Standards and Goals made recommendations that all police departments use formal procedures for screening applicants. The recommendations included the use of measures to assess the personality of applicants. Research in this area increased following this, as did the use of such measures.

3.1 Criticism of Personality Test Use in Police Selection

While the use of psychological tests has become widespread, there has been a considerable amount of controversy as to whether testing has any validity for police selection. An example of criticism in the past is Burbeck and Furnham (1985) who cast doubt on selecting police officers using measures of psychological characteristics. They found no test that was able to distinguish consistently between those who will be successful police officers and those who will not. They state that the test profiles of police are almost always within the normal range, and that using test scores too rigorously will lead to the rejection of suitable officers.

Burbeck and Furnham also refer to the issue of the predispositional and socialisation models. That is, if there is a prototypic ideal police personality and applicants are selected to meet these characteristics, their personality may change over time after selection. If experience on the job develops undesirable characteristics, assessing characteristics at the time of selection may not be unhelpful.

Burkhart (1980) places an emphasis on the special circumstances of the work environment in determining how the new police officer adapts. He argues that the relationship between particular work environments and behaviour will be strong, making the performance of the new officer a result of complex interactions between pre-existing personality characteristics and the social and organisational structure of the agency.

Although certain personality traits present in applicants may alter for better or worse during their employment as police officers, this does not seem to invalidate the use of psychological testing. At the least, Police management should be able to know that a new police officer begins his/her career relatively free from psychopathology. It would also seem to be advantageous that new police officers possessed personal qualities that maximised their chance of a successful career. It may be true that positive attributes may change, or that psychopathology may develop, but the environment that creates these negative changes in officers is a separate issue from selection.

Burbeck and Furnham (1985, p. 68) concluded on the basis of available research that "psychological testing is an unreliable and unpredictive method of selecting candidates to a police force". They conceded that tests might be of use in screening out those with mental problems of some description, although subjects may be able to fake good successfully.

Despite the issues raised by Burbeck and Furnham and others in the past, the use of personality testing has become widely accepted. In fact, countries such as the United States have made psychological testing mandatory for police

departments. A failure to adequately screen police applicants may lead to litigation for negligent hiring if an officer's psychological problems lead to harm to a member of the public (Gaines, Kappeler and Vaughn, 1997).

More recent developments in the area of personality assessment have brought increased confidence about predicting performance using normal psychological characteristics (Barrick & Mount, 1991). The five factor model of personality, or "Big Five" has become the foundation for much research by providing a taxonomy of personality attributes. This model of personality has become a dominant force in personality psychology, and here personality can be understood in terms of five dimensions: Neuroticism vs Emotional Stability, Extraversion or Surgency, Openness to Experience or Intellect, Agreeableness vs Antagonism and Conscientiousness or Will to Achieve (Costa & McCrae, 1992b).

Cortina et al. (1992), as well as Barrick and Mount (1991) identify the development of the five factor model as being pivotal in establishing correlations between personality and performance. However not all view the Five Factor model as being the best model for conceptualising personality, and there are numerous critics (for example Butcher & Rouse, 1996, Ben-Porath & Waller, 1992).

The meta-analysis conducted by Barrick and Mount (1991) using this taxonomy of the Big Five did show consistent relationships between personality attributes and work performance across a number of occupations including police. They identified Conscientiousness and Emotional Stability (or Neuroticism) to be most important to work performance. Conscientiousness assesses characteristics that are important and valuable for any job. These characteristics include persistence, responsibility and diligence. Emotional Stability is also thought to be important across all occupations, as neurotic characteristics such as anxiety are likely to interfere with work performance.

Cortina et al. (1992) also identified Neuroticism and Conscientiousness as being the best predictors in a police sample across a variety of criteria, such as ratings by supervisors and turnover. However, the personality factors added little to the predictive validity of the Civil Service exam, which is primarily a measure of cognitive ability.

3.2 Screen in/Screen out

"Screening in" refers to the selection of personnel by assessing desirable characteristics, while "screening out" means selection is based on excluding those with undesirable characteristics. Some controversy has been generated as to whether personality testing should be used to screen in those applicants who are best suited to a police career, or screen out those who possess undesirable psychological characteristics (McQuilkin, Russell, Frost, & Faust, 1990).

A case may be made for selection of police on either desirable or undesirable characteristics depending on the base rate of psychopathology assumed in the applicant population. Saxe & Reiser (1976) have emphasised the importance of looking at the variations in normal personality characteristics when selecting police officers, because of the low incidence of psychopathology found in this population. Using the MMPI, Carpenter and Raza (1987) found that police applicants as a group were more psychologically healthy than the normal population. They were less likely to be depressed or anxious, and more likely to be assertive and sociable. There were few applicants with significant clinical elevations, but as a group they were more likely to be trying to present a good impression. However, Lorr and Strack (1994) found one in four police applicants to have relatively high levels of paranoia, schizophrenia, and other pathological symptoms.

There are certain characteristics that are generally agreed to render a person unsuitable for work as a police officer. These include psychopathological

disorders and emotional instability. Anxiety, impulsiveness, misperception, low assertiveness and withdrawal have all been identified as being detrimental to the work of a police officer (Hargrave & Berner, 1984, cited in Shusman, Inwald, & Knatz, 1987).

In the past, psychological testing was seen as being able to more accurately identify those with negative characteristics, rather than those who possess positive attributes (Bennett, 1990; Blau, 1994; Burbeck & Furnham, 1985; Meier, Farmer, & Maxwell, 1987). Meier et al. (1987) make a case for the primary importance of screening out unsuitable individuals. They point out that screening in relies on the ability of instruments to make predictions about future behaviour that have been shown to be less than ideal. Screening out, however, utilises the ability of instruments to assess current behaviour. In light of the serious nature of police work, it would appear prudent to use the best possible means of screening out those who are unsuitable before attempting to select those who may be the best applicants.

Selecting the "best" candidates for a position would seem to be the most efficient method of selection, as this process might also then eliminate unsuitable applicants. However, an applicant who scores highly on a desirable aspect of personality such as conscientiousness may also possess undesirable qualities such as anxiety or compulsiveness.

Henderson (1979) and Beutler, Storm, Kirkish, Scogin, and Gaines (1985) describe the difficulty of measuring the success or otherwise of screening out unsuitable personnel. The central problem is the difficulty in establishing criteria by which to measure the predictive validity of the selection program. Many of the inappropriate behaviours (such as abuse of lethal force) among police officers, which are hoped to be avoided by careful screening, occur at a very low frequency. They also often occur in unusual situations that are not normally encountered.

Hiatt and Hargrave (1988) examined the job performance of a group of police officers of whom some had been employed against the recommendations of the screening psychologist. Eleven of the 15 officers who were assessed as being unsuitable were rated as unsatisfactory on the job. However, 13 of the 40 officers who were evaluated as being suitable were subsequently rated as having unsatisfactory work performance. This was expected because of the range of other factors beside psychological characteristics that might produce work problems. The study provided some evidence of the ability of a psychological evaluation to screen out those would later perform poorly as a police officer. However, the high number of officers evaluated as suitable who performed unsatisfactorily showed that the screening in function of the evaluation was less successful.

Fyfe et al. (1997) describe the need to screen out psychologically unfit applicants as a critical requirement of selection. Fyfe et al. believe that because of the low predictive value of psychological tests, they are best utilised to screen out the extreme cases rather than being relied on to decide who is most suitable for police work (screening in).

Hogg and Wilson (1995) discuss the consequences of selection errors based on both approaches to screening. When screening out, a person who may prove to be a good police officer may be incorrectly rejected. When screening in, it is possible that an unsuitable officer may be accepted. The risks and costs of the latter error are the most severe (Rowe, 1985, cited in Hogg & Wilson, 1995).

However, while psychological evaluation to weed out the worst of the potential applicants will help to ensure that those selected meet a minimum standard, it does not assist in choosing who may be the best of the remaining applicants. This is one reason for the increasing popularity of tests that measure aspects of normal personality, and can be used to assist in screening in those most suited to police work.

Tests such as the NEO-PI-R have been developed to measure the "Big Five" dimensions of personality. The strength of tests such as the NEO-PI-R in selection seems to lie in the assessment of stable personality characteristics that are to some extent predictive of performance. Because they measure normal personality characteristics, they may be less than ideal for evaluating psychopathology when used on their own (Ben-Porath & Waller, 1992).

Ideally selection will perform both screen in and screen out functions (Hogg & Wilson, 1995). Screening in the best available applicants may be the preferable approach. The development of the five factor taxonomy, and the consistent relationships found between personality and performance has led to screening in becoming a more realistic option. However, despite this advance, predictive validity remains relatively low (Fyfe et al. 1997).

The assessment of current psychological health (corroborated by other sources of information) in order to detect existing problems seems to remain a priority. This assessment requires psychological tests designed to identify psychopathology (such as the PAI) rather than tests that assess aspects of normal personality (such as the NEO-PI-R).

3.3 New Zealand Police Selection

Until 1995, psychological testing in the New Zealand Police incorporated a cognitive ability test and the Police Entry Test, which assesses work related skills such as observation and report writing. In 1995, Black evaluated the Revised NEO Personality Inventory (NEO-PI-R) as a means of assessing personality characteristics relevant to a career in the police.

Black (1995) set out the selection criteria used by the New Zealand Police to evaluate the suitability of a personality test for recruitment purposes. These are described below in random order:

1. Normal vs. Abnormal: A test which assesses dimensions of normal personality was thought to be of greater utility for selection by allowing strengths and weaknesses (aside from psychopathology) relevant to police work to be evaluated.
2. Validity and Reliability: A high degree of test-retest reliability and criterion-related validity.
3. Scales: Scales must include neuroticism/emotional stability, extroversion/introversion and conscientiousness.
4. Acceptability: Easy to understand, attractive in presentation and based on published empirical evidence.
5. Fairness: Equitable to all ethnic groups and both genders
6. Cost: Low cost per applicant for materials.
7. Scoring: Test must be able to be scored and interpreted by Psychological Services in 5 days.
8. Availability: Supplied by the Test Library, New Zealand Council for Educational Research.
9. Applicability and Administration: Includes scales which can be used for selection and assessment at recruit and executive levels and preferably supplying information useful for counselling. Easy to administer to groups.
10. Time: Able to be completed in less than one hour.

The criteria listed as 1 and 3 appear to be based on a screening in approach. It utilises an assessment of normal aspects of personality such as conscientiousness, which have been shown to predict performance (Barrick & Mount, 1991; Black, 1995). These criteria may not be adequate for several reasons.

As has been discussed previously, it may be important to include the screening out of unsuitable individuals as a primary goal. In the context of personality testing in police selection, unsuitable individuals are those who fail to meet standards of psychological fitness. The standard of the New Zealand Police (1998, p. 70) includes "the absence of, or a minimal presence of, the following: major psychological disorders, anxiety, depression, stress disorders, post-

traumatic reactions, serious substance abuse, disordered interpersonal relationships, disorders of impulse control, excessive aggression, and psychosomatic disorders".

The standard of psychological fitness set by the New Zealand Police cannot be assessed using an instrument based on Black's (1995) criteria. The criteria given for test selection exclude tests that assess for psychopathology, in favour of tests based on an assessment of normal personality characteristics.

It may also be inappropriate to use just one personality test in isolation (Aylward, 1985). The need to both screen out pathology and screen in desirable characteristics may be asking too much of a single instrument. In addition, the issue of defensive responding has not been addressed and may threaten the validity of any psychometric assessment used (Ben-Porath & Waller, 1992).

As a measure of normal personality, the NEO-PI-R is not focused on detecting psychopathology, and may not be sufficiently sensitive to undesirable psychological characteristics such as personality disorders and substance abuse tendencies. However, other screening methods such as background checks and interviews are also used, and these have been thought to identify most unsuitable applicants.

Despite some possible shortcomings, the NEO-PI-R has proven satisfactory to the New Zealand Police in terms of its ability to measure desirable characteristics such as conscientiousness to assist in making selection decisions. Black's (1995) study showed that the NEO-PI-R demonstrated adequate validity for predicting recruit performance, and added significant incremental validity when used in conjunction with the existing measures of cognitive ability and job related skills.

However, there have been concerns that a few individuals with some psychological difficulties have been selected, and have not been identified until

they are undergoing training (I. Miller, personal communication, March. 1999). It is a difficult process to end a recruit's employment as this stage, and also represents a waste of resources. This has led to consideration of instruments to fit a narrower role than that of the NEO-PI-R in specifically identifying psychopathology.

The new instrument is to be used as an adjunct to the NEO-PI-R, and administered to applicants about whom recruiters have concerns as part of a process of further investigation. The NEO-PI-R may be used to "flag" individuals requiring this further investigation, as is being done by the Queensland Police Service with the 16PF (Police Education Advisory Council, 1998). The test is also to be used as a standard part of the screening process for specialist police units such as the Armed Offenders Squad, and the Special Tactics Group who perform anti-terrorist functions. At the present time, the Personality Assessment Inventory (PAI) is being considered to fit this role.

Blau (1994) in a review of personality tests for police selection commented the PAI has promise as a more contemporary measure of relevant psychological issues, as it is based on DSM-IV diagnostic criteria. The use of modern diagnostic constructs gives the test more clinical relevance than the MMPI, which has been the predominant measure of psychopathology (White, 1996).

Chapter 4: Review of Tests Used in This Study

4.0 PAI

The PAI was developed by Leslie Morey and published in 1991. The PAI has appeared to gain considerable popularity in recent years. Surveys during the early 1990s did not find the PAI in the top 30 tests used by clinicians for assessment (Watkins, Campbell, Nieberding, & Hallmark, 1995). However, more recently, it has been ranked in the top 10 tests used by forensic psychologists for cases involving emotional injury (Boccaccini & Brodsky, 1999). Piotowski and Belter (1999) found the PAI to be ranked fourth by directors of internship training in terms of objective personality tests used by interns. Another survey (Belter & Piotrowski, in press, cited by Piotrowski, 2000) found the PAI to also be ranked fourth in assessment training in clinical psychology programs, behind the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, 1994), and NEO-PI-R in terms of emphasis in course work.

Morey's original conception of the PAI was based in part on Jackson's (1971) arguments for rational item selection practices. Jackson argued that the most efficient way to measure a trait was to assess the behaviours most relevant to the trait. This is as opposed to empirical methods (such as those used in the development of the MMPI), which gather a group of heterogeneous items, and empirically derive items that differentiate between groups of interest. These items do not necessarily bear any obvious relevance to the construct being measured.

The PAI is self administered and consists of 344 items that may be answered on a Likert-type scale with the anchors Totally False, Slightly True, Mainly True, and Very True. Morey (1996) sees the four point alternative as having merit psychometrically by allowing more variance to be measured per item, and

also clinically as it allows clients to express their perception of themselves more accurately. Clients have been unhappy with the true false dichotomy found in many tests when they feel that they fall in between these two poles.

The PAI items were selected using a number of parameters based around both the conceptual nature (such as an adequate coverage of the targeted construct) and the psychometric adequacy of the item. Morey (1996) uses the example of items included in each PAI scale that measure the full range of severity of a particular disorder or construct. This is not possible when test items are chosen based solely on discrimination between groups.

The PAI uses validity scales such as PIM (Positive Impression Management) and profile configuration information in order to detect defensive responding. However, as Morey (1996) points out, while these strategies may provide valuable information, they are not infallible and detecting defensive responding remains difficult.

The effects of age, gender and ethnicity need to be considered when evaluating the PAI in the New Zealand Police context. The influence of demographics has been found to be low with a few exceptions (Morey, 1996). The only scale for which ethnicity is associated with scoring is for Paranoia with non-whites scoring six higher on *T* scores. Age has an influence on five scales with effects ranging from minus four *T* to plus seven *T*. Those who are 29 years old or less score five *T* points higher on PAR (Paranoia), six *T* points higher on BOR (Borderline Features), seven *T* points higher on ANT (Antisocial Features), five *T* points higher on AGG (Aggression), and four *T* points higher on STR (Stress). Gender affected only one scale (Antisocial Features) with males scoring three *T* points higher.

The PAI has been identified by authors such as Helmes (1993) and Rogers, Ornduff, and Sewell (1993) as having a number of advantages over other objective measures of psychopathology such as the MMPI-2, and MCMI-III. The four point Likert-type scale allows for the capture of more variability than

creates the potential for enhanced discriminative validity. The MMPI-2 and MCMI-III have items loading on to more than one scale, allowing longer scales and hence increased reliability. However, this sacrifices some discriminative validity, as it forces a correlation between measures of what should be distinct constructs (Retzlaf & Gilbertini, 1987).

However, Boyle and Lennon (1994; 1996) have had some criticism of the PAI. Their concerns seemed to be primarily the potential offensiveness of some test items, reliability issues, age effects, discriminative ability and questions about the factor structure of the PAI. Morey (1995) replied to their critique, and Conger and Conger (1996) also gave their perspective on the issues raised.

Boyle and Lennon (1994) cite several test items that they believe may be considered offensive by some. These are "I have little interest in sex" (measuring a physiological aspect of depression) and "I frequently have diarrhea" (measuring an aspect of somatization disorder). They cite the personal and intrusive nature of such items as being potentially problematic. It seems debatable however whether these items are any more intrusive or sensitive than other items such as those that ask about psychotic experiences for instance. Questions about interest in sex are a common feature of tests including well-accepted measures such as the Beck Depression Inventory – Second Edition (BDI-II; Beck, Steer, & Brown, 1996).

Boyle and Lennon found test-retest reliability to be lower than expected with a median stability coefficient of .7 with a 28 day period between administrations. They also found the median alpha coefficient across full scales to be .83, a figure they considered to be too high and suggesting the presence of redundant items. It is not clear on what basis they make this judgement. Kline (1993) asserts that the majority of psychometrists give high internal consistency as a prerequisite for a good test

Boyle and Lennon found significant differences between age groups and gender on the Schizophrenia and Alcohol Problems scales. Morey (1995)

points out a confounding element that is likely to have produced this effect. Boyle and Lennon compared young normal college students with a sample of middle aged clinical subjects, and so the presence of age differences on clinical measures would be expected.

The PAI was unable to differentiate between a group of alcoholics and schizophrenics on the Schizophrenia scale. However, as Morey (1995) pointed out, the mean score on the Negative Impression Management scale for the alcoholic group was well in excess of a score suggesting an exaggeration of pathological symptoms. This may have produced an elevation on the schizophrenic measure. In addition, the sample of schizophrenics were hospitalised and on medication which may have reduced their test scores.

The PAI Alcoholism scale identified 25% of the normal group as meeting criteria for alcohol abuse. Boyle and Lennon suggested that this demonstrated a high rate of false positives as they quoted a base rate of abuse as .18%. They later admitted this should be 18%, but based their arguments on the lower figure (Conger & Conger, 1996). The magnitude of the error in this case not only invalidated their argument on the issue of a high false positive rate, but also detracted from the credibility of the study.

The discussion around the factor structure of the PAI was based in part on the technique used to extract and rotate factors. Boyle and Lennon produced a different factor solution from that of Morey (1991). However, a different factor analytic technique was used on a different sample, and Conger and Conger (1996) comment that differing factor solutions would be expected in this situation. The differences found did not show which one might be more accurate and did little to shed light on this issue.

Conger and Conger summarised the debate raised by Boyle and Lennon as raising important issues, but containing significant problems. These were identified as confounds existing between sample groups, an incorrect alcohol abuse rate, and overstated concerns about the psychometric qualities of the

PAI. It was difficult to draw any conclusions from the study because of the methodological flaws present. However, it did highlight the need for caution when evaluating a test based on small samples that differ considerably from the normative population group.

The PAI has been used in published studies investigating reliability and validity in a number of populations, including police, substance users, prison inmates and psychiatric inpatients. Hays (1997) compared the PAI and MMPI profiles of nine police applicants, and concluded that the PAI demonstrated concurrent validity. However, the small sample size and the use of police applicants (rather than recruits) make comparisons of Hay's data with the present study difficult.

Rogers, Ustad, & Salekin (1998) found the PAI to demonstrate moderate to high convergent validity with the Schedule for Affective Disorders, Structured Interview of Reported Symptoms, and the Suicide Probability Scale in a sample of prison inmates. They concluded that the PAI demonstrated considerable clinical utility for identifying psychopathology in this population. Wang, Rogers, Giles, Diamond, Herrington-Wang, and Taylor (1997) also found the PAI to demonstrate high convergent validity in a corrections setting.

Boone's (1998) study investigating internal consistency reliability in a psychiatric inpatient group found the full scales to have high coefficient alpha levels averaging .82, with reasonably low standard errors of measurement. Subscales had lower reliability averaging .66 with higher standard errors. He concluded that full scales could be interpreted with some confidence. However, subscales needed to be treated with some caution, and used more for generating hypotheses for further investigation rather than for diagnosis.

Many of the published studies of the PAI have focused on the use of validity scales to detect response sets, and defensiveness in particular (eg. Baer & Wetter, 1997; Fals-Stewart & Lucente, 1997; Fals-Stewart, 1996; Morey, 1991; Morey & Lanier, 1998; Cashel, Rogers, Sewell, & Martin-Cannici, 1995;

Peebles & Moore, 1998). The studies have found varying degrees of success in using the validity scales to detect response sets.

Cashel et al. (1995) found the PIM (Positive Impression Management) scale correctly identified just under half of the participants instructed to present a favourable impression, but correctly identified 84.9 % of those answering honestly. Morey had more success in detecting defensive respondents (81.8 %), but was less accurate in classifying honest respondents (69.6 %). Cashel et al. attributed this in part to differing test instructions given.

Looking at the identification of malingerers, Morey (1991) found that 88.6% of those simulating disorders were able to be identified with the NIM (Negative Impression Management) scale, with only 10% of the genuine clinical participants misclassified. However, Rogers, Ornduff, and Sewell (1993) had less satisfactory results. They found that while detecting those feigning schizophrenia was successful (90.9%), only 55.9 % of those feigning depression and 38.7 % of those feigning generalised anxiety disorder (38.7%) were identified. Rogers, Sewell, Morey, and Ustad (1997) subsequently developed a discriminant function that resulted in an 80% accurate identification of those feigning either schizophrenia, depression, or generalised anxiety disorder.

The literature available on the PAI presents it as a reliable, valid instrument for the assessment of psychopathology. The method of test construction used gives it some significant advantages over many other instruments in providing superior psychometric qualities, and allowing test scores to be easily interpreted by clinicians. However, the obvious nature of items leave it opened to being faked by respondents. The validity scales included in the PAI to assess response distortion have been shown to be relatively effective, but their accuracy is less than ideal.

4.1 NEO-PI-R

Published in 1985 by Costa and McCrae and revised in 1991, the NEO-PI-R in its present form is based around the "Big Five" dimensions of personality. The NEO-PI-R was introduced in 1995 as a standard part of assessment of police applicants and is in use at the present time. Black's (1995) trial of the NEO-PI-R for the New Zealand Police found an uncorrected correlation of 0.38 between the five personality factors and recruit performance during training, comparable to that found for the mental abilities tests already in use. Significant incremental validity when combined with a test of cognitive abilities and a job skills related test was also demonstrated with the correlation between scores and performance increasing to 0.50. This was cited as evidence of the NEO-PI-R's utility in selecting suitable police recruits.

The NEO-PI-R has been criticised for its lack of validity scales. There is only one validity item that asks if the person taking the test has responded honestly. This may leave the test open to misrepresentation by the respondent in either an overt or subtle manner (Ben-Porath & Waller, 1992). Although the NEO-PI-R may not be unduly affected by socially desirable responding when respondents are cooperative (Costa & McCrae, 1992b), the same cannot be assumed under special circumstances such as a job application. Those completing the NEO-PI-R under these circumstances may have a high degree of motivation to respond in a socially desirable manner as has been found with other instruments such as the MMPI (Carpenter & Raza, 1987; Hays, 1997). The issue of response distortion is covered in more detail in Chapter 5.

Chapter 5: Special Issues in Test Evaluation

5.0 Overview

This chapter looks at some aspects of test evaluation that are particularly relevant to the context of personnel selection. A basic discussion of psychometrics has been omitted in favour of an examination of those issues that are particularly problematic in personnel selection. The first two sections look at the problems involved in assessing criterion-related validity, while subsequent sections of this chapter examine the issues of response distortion and test fairness.

5.1 Criterion-Prediction Validity

This process measures the ability of a test to predict the respondent's performance in particular activities. It is most pertinent when tests are being used to select personnel, and helps to assess the utility of a test for a particular selection application. Hogg and Wilson (1995) state that for a test to be valid for selection, it must predict one or more job behaviours.

Difficulties in using criterion-prediction validity have been found where sample sizes are small, resulting in correlations between test scores and performance failing to reach statistical significance. Schmidt, Hunter and Urry (1976, cited in Anastasi and Urbina, 1997) found that half of the samples used to validate tests in industrial settings had 50 cases or less and that criterion-related procedures were not applicable to such small samples. Use of meta-analytical techniques has been able to overcome this problem and have revealed significant relationships between test measures and performance in the workplace.

A significant problem identified in police selection research is finding adequate criteria against which to measure the predictive validity of a test (Bartol, 1991).

The most commonly used criterion has been supervisory ratings. Use of this measure has been troubled by difficulties with a lack of job-relevant criteria (Coutts, 1990), and poor interrater reliability (Rothstein, 1990).

Much of a police officer's shift is spent unsupervised, as he/she is only accompanied by a fellow officer rather than someone in a supervisory position. Any rating then cannot be based on a full assessment of the officer's work performance, as the supervisor is only in a position to observe a small part of it (Hargrave & Hiatt, 1987).

The use of measures during training has been another common criterion used, and has the advantage of being more standardised than job-related criteria (Hogg & Wilson, 1995). However, the relationship between training success and on-the-job performance has been generally shown to be poor (Kleinman & Gordon, 1986). The psychological characteristics required to succeed in what is primarily an academic environment may be different from those required in the workplace (Hiatt & Hargrave, 1988).

In the present study, criterion-related validity is considered to be of interest, but secondary to the goal of assessing the participant's current psychological health. Instruments such as the PAI have been typically found to demonstrate low levels of predictive validity, as they measure abnormal characteristics that are uncommon in highly selected police samples. Although the frequency of abnormal psychological characteristics may be low, the nature of police work makes their detection important.

When using a measure of psychopathology for selection, job related behaviour might not be a particularly useful criterion for assessing the value of a test. The frequency of negative behaviours such as misuse of force, or the consequence of these behaviours such as complaints from the public is so low as to make them unhelpful as performance measures (Henderson, 1979). What may be more important to demonstrate is concurrent validity, in that negative

psychological characteristics indicated by test scores are corroborated by other information such as clinical interviews.

In a measure of psychopathology, evidence of concurrent validity may also be demonstrated through its correlation with other instruments measuring the same constructs. In the case of the PAI, this has been demonstrated through strong correlations with measures such as the Beck Depression Inventory and the MMPI (Morey, 1991).

Range restriction is a common problem in research where recruits or serving police officers are used as participants. The selection that takes place during the recruitment procedure leads to a restriction in test scores which can produce an underestimate of the true relationship between scores and performance (Hogg & Wilson, 1995). This is a particular issue in the current study as the recruits have been through a comprehensive selection progress including completing the NEO-PI-R.

There does not appear to be a simple answer to the problem of range restriction causing underestimates of predictive validity. Developing a database of applicant scores and relating this to scores of incumbents may assist in developing more accurate assumptions on which to base correction formulas. In the meantime, evaluation of predictive coefficients needs to include an awareness of the likelihood of underestimation.

5.2 Response Distortion

When interpreting the scores derived from a psychological test, a first step must be a consideration of the quality of the data that has been gathered. It is critical to evaluate the respondents' approach to completing the test when there is apparent motivation to distort responses in a positive or negative direction (Ben-Porath & Waller, 1992).

A variety of sources identify the distortions present in self report data as a significant threat to the validity of personality tests particularly in selection situations (Burbeck & Furnham, 1984; Carpenter & Raza, 1987; Henderson, 1979; Hays, 1997). Tests such as the NEO-PI-R are commonly used in clinical situations to assist in providing the most suitable treatment to clients (Costa & McCrae, 1992b). In this situation, there appears to an incentive to the client to respond honestly as the test may be viewed as part of the process of gaining relief from what is troubling the person. However, the police applicant presented with a test which has the stated purpose of assessing his/her suitability for police work seems to have every motivation to provide at the very least guarded responses and possibly even distortions. Costa and McCrae (1992b) comment that where there are strong incentives to distort self-presentation, self report data is unlikely to be reliable.

Hough, Eaton, Dunnette, Kamp, & McCloy (1990) found that distortion of responses in applicant and incumbent military samples did not significantly affect criterion-related concurrent validity. The predictive validity coefficients for a group of soldiers scoring high on a measure of socially desirable responding were compared with those scoring lower. Little or no difference in predictive validity was found between the groups. They concluded that intentional distortion was not a serious problem for this population, although they did recommend that other sources of information be used to make selection decisions about applicants who had been identified as distorting responses.

Rosse, Stecher, Miller, and Levin (1998) believe that there are many reasons why the suppressing effects of response distortion on predictive validity have not been identified in previous studies. The first relates to the nature of socially desirable response distortion in that it is thought to have a two-factor structure. The first factor represents an unrealistically positive view of self and is correlated with measures of positive adjustment. The second factor is the deliberate faking of answers to create a positive impression. Paulhus and Reid (1991) found that scores on the second factor were far more sensitive to

situational demands than for the first factor. Rosse et al. (1998) comments that most measures of response distortion load on both factors.

Rosse et al. cites other reasons for the apparent lack of impact of response distortion on predictive validity. These are the skewed distribution of response distortion, range restriction, and the low predictive validity of personality tests. They note that while response distortion may have no discernible effect on predictive validity, it can have a dramatic effect on hiring decisions. If a small proportion of applicants extremely distort their responses, their scores may place them ahead of applicants who have been more honest in their completion of the test.

Topping and O'Gorman (1997) examined the effects of a faking set on the validity of the NEO-FFI (a briefer version of the NEO-PI-R), and found that the scores of those instructed to present positively changed significantly in the direction expected. Means for the agreeableness, conscientiousness, and extraversion scales were higher, and the mean for neuroticism was lower than for the honest condition. The self report scores were compared to judges ratings using the observer version (where another person completes the test based on observations of the subject) of the test in order to demonstrate changes in validity. All scales in the faking condition except agreeableness showed statistically significant decreases in validity compared to the honest condition.

Ross, Bailley, and Millis (1997) found that the NEO-PI-R was highly sensitive to positive self-presentation effects. The scales of Agreeableness and Conscientiousness were thought to be most easily affected by positive self-presentation. This is of concern, given the importance of Conscientiousness in predicting work performance (Barrick & Mount, 1991).

The issue of defensive responding when testing law enforcement personnel does not appear to be straightforward. While Morey (1991, 1996) and others (Cashel et al., 1995; Peebles & Moore, 1997) have reported relatively low

scores on the PAI defensiveness scale (PIM; Positive Impression Management) as being indicative of defensive responding, Roberts, Thompson & Johnson (1999) assert that these scores are misleading when applied to a law enforcement sample. They state that the contextual demands of the testing situation produce elevations in validity scales when using community norms, and that the validity of a protocol should not be judged on this basis. They describe one case of a police applicant with a highly elevated PIM (Positive Impression Management) *T* score of 70 (using community norms) who showed no sign of defensiveness or evasiveness when interviewed and assessed by other means including polygraph examination. This was described as being a typical finding when interpreting the PAI profile of law enforcement personnel.

As well as characteristically elevated PIM scores, Roberts et al. (1999) report that public safety personnel typically have suppressed *T* scores on clinical scales using community norms. This has resulted in some difficulty in making interpretive decisions, as almost all respondents have scores below clinical cutoffs. Their response has been to produce norms using a large law enforcement and public safety sample (N = 17757), and to make screening decisions based on this data. The usually low level of endorsement of clinical items then creates a large spike in *T* scores for those respondents who do endorse such items. This aids in identifying those who differ considerably from the law enforcement sample. They do stress that interpretive decisions about identifying actual pathology should be based on community norms.

While the large database compiled by Roberts et al. lends some weight to his conclusions, the apparent casting aside of research relating to defensiveness particularly for the PAI, is possibly of some concern. It is not clear why a level of defensiveness thought to be indicative of less than honest responding in a community sample should be thought to be acceptable in a law enforcement sample. Both applicants and serving officers would seem to have every motivation to present as positively as possible, and to minimise any psychological difficulties experienced.

The researcher was unable to access the data presented to support the view of Roberts et al., as it has not been published. Distortion will occur in an applicant sample (Hough, 1998), and a "common sense" approach would suggest that a motivated applicant would be inclined to respond negatively to obvious items such as "Sometimes I am very violent". It seems premature to dismiss the impact of response distortion in a population with every reason to present well.

Costello, Schneider and Schoenfield (1993) found that elevated K scores (defensiveness) on the MMPI were positively correlated with disciplinary suspension days in a small sample of police applicants. They identify defensive responding as a significant threat to the validity of self report tests in a selection context.

Roberts et al. (1999) appear to have used a variety of other measures and interviewing to cross validate conclusions drawn from PAI scores. Roberts in his practice has found that these other sources of information have not usually revealed evasiveness or pathology even when PIM scores are highly elevated (against community norms). Hogan, Hogan, & Roberts (1996) suggest that being able to distort scores in a socially desirable manner may relate to increased job performance because of an awareness of social norms. The perceptiveness and interpersonal skills necessary to perform police duties may serve an applicant or police officer well when positively dissimulating on tests or interviews.

Roberts et al. (1999) cite the finding of Butcher, Morfitt, Rouse and Holden (1999) who found that 27% of applicants for airline pilot positions produced invalid profiles on the MMPI-2. Roberts et al. present this as evidence of the normality of elevated validity scales in an applicant sample, rather than as cautionary evidence about the applicability of self-report testing. An interesting finding relating to the study of Butcher et al. is what happened when the pilot applicants were retested, after being told their first test had been detected as being defensive. They tended to produce valid profiles on the second administration and in 12% of the cases, scores moved into the clinical range.

The implication of findings such as this are interesting when applied to the selection of police where similar contextual demands seem likely to apply in the testing situation. Defensive responding may be suppressing scores on clinical scales to an extent that some individuals who are experiencing some psychological difficulties are able to escape detection. However, low scores on clinical scales may be true scores as asserted by Roberts et al., and elevated defensiveness scores are a function of the specific population and testing situation.

It is difficult to know if giving warning of the ability of the PAI to detect defensive responding prior to administration would actually reduce defensiveness. Although this appeared effective in the Butcher et al. study when given to those who had been identified as defensive on a first administration, the effect of having being detected may have provided additional incentive for those respondents to then answer honestly.

Hough et al. (1990) cite four studies (Haymaker & Erwin, 1980; Lautenschlager & Atwater, 1986; Schrader & Osburn, 1977; Trent et al., 1986) which looked at the effect of warning respondents about the presence of methods to detect response distortion, and the consequences of doing so. All studies found that the warnings reduced the amount of intentional distortion. What Hough does not mention, however, is that warnings may simply serve to produce more believable distortion. Respondents could still hide negative characteristics while being careful not to overly endorse positive ones. Baer and Wetter (1997) found that providing information about the presence of validity scales resulted in increased ability to fake good without elevating the PIM (Positive Impression Management) scale of the PAI.

The central issues in the debate on defensive responding are it's possible effects on predictive validity and diagnostic accuracy. The work of Hough et al. (1990) in particular is often cited as evidence that defensive responding is of little concern when making selection decisions because of its lack of impact on

criterion-related validity. However, Rosse et al. (1998) have identified some reasons for the effects of defensive responding failing to be detected. In addition, when using an instrument intended to assess psychopathology such as the PAI, defensive responding remains a concern in that applicants with undesirable psychological characteristics may easily reduce their scores to acceptable levels.

5.3 Test Fairness

The New Zealand Police have a commitment to working toward a police service whose composition is representative of the general population (Black, 1995). Hogg and Wilson (1995) state that proportional representation will assist in improving relations between Police and ethnic groups, reduce the "us versus them" syndrome, improve the ability of the Police to understand minority groups, provide potential applicants from minority groups with role models, and improve the credibility of the Police as an agency serving the community.

The representation of minority groups and women in the New Zealand Police has been seen as a priority, and this target has been addressed through a promotion of the Police as a desirable career. This is an alternative to other approaches such as affirmative action where quotas are set, and differing selection standards may be used to fill these (Black, 1995).

Although standards are applied to applicants without regard for their demographics in order to select the best personnel, a concern remains that characteristics of assessment criteria may be biased against certain groups. This may be particularly pertinent to personality tests, where it has been argued that test items may be culturally biased and discriminatory against those of low socio-economic status. This may be a function of wording specific to the majority group and norms based on a white population (Winters, 1988, cited in Hogg & Wilson, 1996). However, DuBois, Sackett, Zedeck, and Fogli

(1993) found that differences were often found between European and African Americans on cognitive tests, but rarely for personality tests.

The construction of a test provides an opportunity to minimise the impact of between-group differences by selecting items that are have been screened for bias against any particular group. The development of the PAI is an example of this process where items were reviewed by a panel that was representative of the community. This panel was asked to identify items that might be offensive to any group and/or might be more likely to be endorsed in a pathological direction as a result of normative beliefs in particular cultures. This review resulted in a number of potential items being dropped from the pool (Morey, 1991).

When evaluating a test, the issue of potential adverse impact on minority groups is important to consider. There are a number of methods for evaluating adverse impact. The first is the four-fifths rule where adverse impact is assumed if the selection rate for any group is less than 80% of the rate for any group with the highest selection rate (United States Equal Employment Opportunity Commission, 1978, cited in Hogg & Wilson, 1996).

Another method is to evaluate a group's job performance level in relation to their test scores. Adverse impact is detected if there is a significant difference between the test scores, but not the performance of a group (Hogg & Wilson, 1996). Evidence may also be drawn from the regression line of a group where predictions based on test scores tend to either over or under predict performance (Maxwell & Arvey, 1993). In this case, the test may still be applicable if different scores are used for differing groups (Arvey & Faley, cited in Hogg & Wilson, 1988).

Arvey & Faley (1988, cited in Hogg & Wilson, 1996) found that tests were equally able to predict job performance for majority and minority groups. The groups did tend to differ on test and job performance measures, however, with differences in test scores being most likely. Using cutoff scores based on the

majority groups' performance would then result in a lower percentage of the minority group being selected.

The discussion of this issue tends to be centered on the ability of a test to predict performance, and may be less relevant when evaluating a clinical measure such as the PAI. The issue that arises from using a test to detect psychopathology is whether differences in responding between groups can be taken as reflecting actual pathology. Most studies in this area have been done with the MMPI, and have found consistent differences in scoring between minority and majority groups (Groth-Marnat, 1997). However, the MMPI has not been found to have greater descriptive accuracy for European Americans compared to African-Americans (Elion & Megargee, 1975; Green & Kelley, 1988). Groth-Marnat (1997) suggests it is possible that differences found may in fact be true differences caused by the increased stresses experienced by minorities. In regard to the MMPI, Graham (1987, cited in Groth-Marnat, 1997) suggests that clinicians should tentatively accept deviant scores, but take particular care to integrate other information with the test scores.

While some effort has been devoted to eliminating bias against any demographic group during the development of the PAI, the test has been constructed using North American samples. Ethnic groups in a New Zealand context may be adversely impacted by a test such as the PAI, or have their scores misinterpreted. While the African American culture may hold some parallels with Maori in being a minority group, this does not mean that both groups will respond in similar ways to a test such as the PAI.

Maori may hold different concepts about mental health and experience psychological difficulties in different ways from other ethnic groups. Durie (1985) discusses how Maori perspectives of health differ from Western ideas of dualism, where the mind and body are considered to function independently. Maori are more likely to take a holistic view of health, and do not separate emotions from the entire bodily experience. One example of this is Maori who are suffering from a depressive disorder may present with different symptoms

from Pakeha. They may be more likely to present with somatic complaints that do not have any apparent physiological cause (National Health Committee, 1996).

The issue of test fairness is a serious consideration for police selection. The literature reviewed found that personality tests are equally able to predict performance for different ethnic groups, but test scores may vary, meaning some groups may be less likely to be selected. This is less of an issue for tests measuring pathology where scores causing concern are likely to be followed up in a full clinical evaluation and background check. This allows the validity of test scores to be checked against observed behaviour and other information.

Chapter 6: The Current Study

This study attempts to evaluate the utility of the PAI as an assessment tool for selecting New Zealand Police. There are two broad goals in using the PAI, firstly to identify applicants who may not meet the New Zealand Police standards of psychological fitness, and second to assist in predicting the performance of recruits and serving officers.

The literature reviewed leads to the conclusion that personality tests need to be used with some caution and scepticism in an applicant setting. Applicants may attempt to hide pathology or exaggerate socially desirable qualities, and personality tests may only possess a modest ability to predict performance.

In the last decade, the five factor model of personality has been shown to reliably predict performance. Factors such as conscientiousness rather unsurprisingly predict good work performance. The New Zealand Police currently use the NEO-PI-R to measure characteristics such as conscientiousness, and utilise this information to screen in applicants with desirable characteristics.

The screening out function of personnel selection may be better met by a test that is designed primarily to measure psychopathology such as the PAI. Without extensive background information and clinical interviews, the diagnostic accuracy of the PAI cannot be ascertained in a study such as this. However, at this stage, it is important that the PAI demonstrate adequate psychometric qualities in the New Zealand police population. This is investigated by measuring internal consistency, and validating the test against the NEO-PI-R.

In an organization that places a high value on equal opportunity, it is important to establish that no applicant is disadvantaged by any demographic

characteristic. This is investigated by looking at the scores of different demographic groups and relating them to performance.

While the literature shows that tests such as the PAI do not possess strong predictive validity in relation to work performance, the PAI is evaluated in this regard in order to make comparisons with the NEO-PI-R. Any evidence of the PAI offering some incremental validity over other tests such as the NEO-PI-R and cognitive abilities tests is of particular interest.

An important issue raised in the review of the literature was the possible impact of defensive responding on the use of personality tests for police selection. While there is some evidence that defensive responding has little effect on predictive validity, it is likely to lower the diagnostic accuracy of a test as a modest reduction in a scale score may place an applicant below the interpretive threshold. The prevalence of defensive responding and its impact on predictive validity is examined.

Evaluation of aspects of the PAI's performance in a New Zealand police population may assist in decision making about the possible introduction of the test, and alert administrators to possible limitations or deficiencies. This evaluation is limited by the absence of corroborating information such as clinical interviews and the reports of others which would normally be utilised when interpreting a test such as the PAI.

Chapter Seven: Methodology

7.0 Participants

The population chosen to be sampled in this study were recruits attending training at the Royal New Zealand Police College (RNZPC). Recruits were used in this study for practical reasons. They are accessible in relatively large numbers in one place and there is a range of data available on them. Each recruit is assessed during his/her time at the RNZPC on a number of measures, and has also completed the NEO-PI-R during their application. There is also a range of demographic information available. In addition, administering the PAI to this group will provide some normative data for this highly selected population. One hundred and sixty recruits were asked to take part in this study with 127 recruits consenting (24 females and 103 males). This gave a participation rate of 79.4%. A summary of demographic details for the sample is presented in Table 1. A summary of demographic details for the entire group is presented in Table 2.

Table 1. *Demographic Details of Police Recruits in the Sample (N = 127).*

	Frequency	Percentage
Age		
20-25	27	21.3
26-30	62	48.8
31-35	21	16.5
36-40	13	10.2
41+	4	3.2
Gender		
female	24	18.9
male	103	81.1
Ethnicity		
Caucasian	110	86.6
Maori	17	13.4
Partner Status		
defacto	16	12.6
engaged	6	4.7
married	43	33.9
single	62	48.8

Table 2. *Demographic Details for all Police Recruits Asked to Participate (N = 160).*

	Frequency	Percentage
Gender		
female	38	24.8
male	122	75.2
Ethnicity		
Caucasian	137	76.1
Maori	23	23.9

The population sampled in this study does not represent either of the populations in which the PAI is likely to be used for assessment (police applicants or serving police officers). However, police recruits were the only personnel available to be tested in relatively large quantities. They do in some ways represent a blend of the target populations having only recently been applicants themselves, and are in the process of training as police officers.

As can be seen from Tables 1 and 2, a higher proportion of Maori and female recruits declined to participate in the present study. However, the percentage of females is still higher than in the New Zealand Police as a whole (New Zealand Police, 1998).

7.0.0 Sample Selection

Dr Ian Miller, the Coordinator of Psychological Services for the New Zealand Police at that time, initially contacted the RNZPC to request the allocation of time for testing. The researcher then liaised with the officer in charge of each training wing to arrange specific times for testing. Wing 186 was the first group of recruits to be tested. There were eighty people in this wing and it was hoped that this wing of recruits would provide a sufficient number for statistical purposes. Participation rates have been very high in the past and it was anticipated that participation would be close to 100% for this study also.

However, participation was not as high as expected with only 56 recruits from Wing 186 taking part. This was thought to be insufficient for statistical purposes and so Wing 187 was also tested.

Each wing of recruits is divided into four sections, and it was arranged to test the recruits in four groups of two sections each. Although it might have been preferable to have smaller groups, the training schedule was full and it was not possible to secure more time for testing. Wing 186 was tested on 5 November 1999 at 1600hrs and 1830hrs. Wing 187 was tested at 1100hrs on 15 November 1999 and 1300hrs on 14 December 1999.

A total of 127 participants took part in this study. Unfortunately, although each participant completed the PAI, NEO-PI-R scores could only be found for 106 of the participants.

7.1 Information Sheet

An information sheet (see Appendix B) was distributed that explained the broad aims of the study, detailed how the data gathered would be used, and explained the limits of confidentiality. It asked permission to use data from the RNZPC, the NEO-PI-R completed during their application, and the PAI. Completion of the PAI test was taken as consent to participate. At the back of the information sheet was included a form which had provision to request feedback of a specific or general nature.

7.2 Test Administration

The researcher visited the RNZPC during the recruits training. Jonathan Black (Manager: Police Psychological Services) attended the first session to supervise the test administration. As the test administration was done to his satisfaction, he did not attend subsequent sessions.

Questions about the study were invited and any concerns discussed. It was made clear to the recruits that participation was entirely voluntary, but their cooperation was requested to assist in this research. Those who did not wish to

participate were able to leave at this point or remain in the room without taking part. The recruits who decided to participate were informed that completion of the PAI implied their consent to participate, and authorised release of other test data to the researcher. Staff from the RNZPC attended only the beginning of the session to ensure that the researcher had any equipment required and left before requests for participation was made.

The procedure for completing the test was explained to the participants at this time and the test materials distributed. No time limit was given for completion of the test. However, the times given for testing did place some pressure on participants to complete the test quickly. The 1100 session was prior to lunch at 1200, the 1300 time slot had a lecture following at 1400, the 1600 session was prior to dinner, and the 1830 session was during the recruits' free time.

The test instructions covered the nature of the item responses and how to fill in the answer sheet. The recruits were asked to answer honestly in order that quality data be collected. No reference was made to the presence of validity scales that might detect particular response sets. The researcher was available during the testing sessions to clarify the wording of test items if required.

7.3 Ethical Issues and Confidentiality

This study conformed with the New Zealand Psychological Society's ethical guidelines. The recruits have been through exhaustive testing during their application to join the Police, and have invested much effort in reaching training at the College. They may understandably have been reluctant to complete a test that may identify them as unsuitable for the police. In order to provide some protection to participants, a guarantee of confidentiality was given. The names of participants would only be used to match their tests with data provided by the College after which time all names would be replaced with a code number. The test scores would not be used for selection or assessment

purposes, and only the researcher would be aware of how each participant scored.

No individual would be identified to Police unless it was believed that the individual presented an imminent danger to himself/herself or others. The decision to disclose an individual's name would be made after consultation with a senior clinician from Massey University.

Each participant was assigned a code number randomly, and all data entered into the computer was referred to using these numbers. All written data was stored in a locked cabinet and computer files were protected with passwords.

7.4 Data Analysis

The following statistics were used to analyse the data collected:-

(1). Descriptive statistics were used to summarise the demographic data and included means, range, and frequencies.

(2). Coefficient alpha and inter-item correlations were used for analysis of reliability.

(3). Pearson correlation coefficients were used to examine the direction and strength of relationships between the tests and performance data.

(4). Independent sample *t*-tests were performed to assess significant differences between the means of different groups.

(5). Regression analyses were carried to assess the predictive validity of the tests.

(6). Principal components analysis was performed to consider the underlying structure of the PAI.

(7). Standard errors of measurement were calculated using internal consistency figures and standard deviations of *T* scores.

7.5 Test Interpretation

The researcher examined the scores of the participants within 48 hours of test administration to screen for item responses that might indicate that the person was a danger to him/herself or others. This was done using the Critical Items Form supplied by the test publisher (Psychological Assessment Resources). It contains twenty-seven PAI items that are considered critical for two reasons: the content of the item may indicate the respondent is in crisis, and normal individuals very seldom endorse these items (Morey, 1991).

The PAI scores were interpreted in accordance with the information provided in Morey (1996) and a brief summary written for each participant. This was mailed to participants along with a summary of the research findings. The interpretive report provided to recruits was restricted to non-clinical facets of the PAI. These were self-concept, interpersonal style and perception of environment. It was felt that providing interpretive information relating to clinical disorders to participants was inappropriate, as it would not be possible to discuss this information in person, or relate it to collateral information.

7.6 Measures

7.6.0 Personality Assessment Inventory (PAI)

The PAI was administered as the primary measure for this study. The PAI is self-administered and consists of 344 items that may be answered on a Likert-type scale with the anchors Totally False, Slightly True, Mainly True and Very True. The items in the PAI are grouped into 22 full scales: four validity, 11 clinical, five treatment consideration, and two interpersonal scales. Ten of the full scales contain subscales, which were intended to assist in interpretation, and to cover complex clinical constructs more fully (Morey, 1996).

Interpreting test scores on the PAI is relatively straightforward with raw scores being converted to *T* scores that are calibrated for each scale. The *T* scores are calibrated so as to have a mean of 50 and a standard deviation of 10 relative to community sample norms. About 84% of nonclinical respondents score below 60 and 98% will score below 70 (Morey, 1996). *T* scores of 70 or more show a prominent deviation from the typical scores of adults living in the community.

Reliability Studies investigating the internal consistency reliability of the PAI have used coefficient alpha. This statistic was used as all items on each PAI scale are intended to measure the same construct. Median alphas for the full scales were .81 for the normative community sample (Morey, 1991). Mean inter-item correlations are also used to measure internal consistency and these are generally about .20. Morey (1991) views this low correlation as positive as it indicates the items are measuring fairly independent content. He seems to feel the broad constructs measured by the PAI would not be adequately sampled by a homogenous set of items.

The temporal stability of the PAI has been investigated by giving the test to individuals on two occasions. For the standardisation study, this gap was four

weeks and the median test-retest reliability for the 11 full clinical scales was found to be .86 (Morey, 1991).

Protocol Validity The PAI contains four scales designed to assess the impact of different response tendencies on test scores. INF (Infrequency) and ICN (Inconsistency) attempt to detect deviations from conscientious responding, while NIM (Negative Impression Management) and PIM (Positive Impression Management) assess attempts by respondents to present an overly negative or positive picture of themselves. There is some difficulty in selecting cutpoints for these scales, as there is a tradeoff between sensitivity and specificity. Sensitivity refers to the ability of the scale to correctly detect those respondents who are positive for the construct of interest such as defensiveness. Specificity refers to the percentage of negatives correctly identified by the measure. As cutoff points increase, sensitivity drops while specificity will increase.

Table 3 presents the decision rules used in this study to assess if protocols are valid. Cutoff scores are based on the recommendations of Morey (1996). The difficulty in establishing one cutpoint that allows validity to be determined with confidence has been addressed by creating a third category beyond valid/invalid, which is "interpret with caution". This range of *T* scores listed as moderate is likely to contain a significant percentage of invalid protocols and interpretations of other scales would be done with some caution. For the purposes of statistical analysis, the "interpret with caution" protocols were included with the invalid ones. This is because only the group classed as valid can be said to be so with some confidence.

Table 3. *Criteria for a Valid PAI Protocol in the Present Study.*

Scale	*T* score	Interpretation
Detecting careless or idiosyncratic responding		
Infrequency (*INF*)		
low	<60	valid
moderate	60-74	interpret with caution
high	>=75	invalid
Inconsistency (*ICN*)		
low	<64	valid
moderate	64-72	interpret with caution
high	>=73	invalid
Malingering		
Negative Impression Management (*NIM*)		
low	<73	valid
moderate	73-91	interpret with caution
high	>91	invalid
Defensiveness		
Positive Impression Management (*PIM*)		
low	<57	valid
moderate	57-66	interpret with caution
high	>66	invalid
Defensiveness Index		
	>70	invalid

For defensive responding, Morey found a PIM *T* score of 57 to be the point above which respondents were more likely to be answering defensively. Using this cutoff score, he found that PIM correctly identified 82% of respondents instructed to answer defensively (sensitivity) and 70% of those answering honestly (specificity). Using 57 *T*, Cashel et al. (1995) found sensitivity and specificity rates of 48% and 81% respectively. They recommend using 57 *T* as a possible cutpoint, and found that a higher cutpoint was poor in detecting defensiveness.

Depending on what data is used, the 57 *T* cutpoint still misclassifies a significant percentage of the samples. For this reason, Morey (1996) uses a higher cutpoint (two standard deviations above the clinical mean) when making more definitive statements about a protocol's validity. Using the two standard deviation rule, Morey (1991) found sensitivity and specificity rates of 86.8% and

92.3% for ICN respectively, 50.5% and 92.5% for ICN, 86.5 and 94.1 for NIM, and 43.1 and 94.9 for PIM. Cashel et al. (1995) found that PIM performed more poorly in their study with sensitivity and specificity rates of 16.5% and 98.6%. They suggest this may have been due to participants in the "fake good" condition having believability stressed to them as part of defensiveness.

The PAI Defensiveness Index (DEF) was also used to detect overly positive self-presentation by respondents. This index is made up of eight configural features that are seen much more often in the PAI profiles of those instructed to present positively. Morey (1996) suggests that DEF be used as a "rule in" tool, rather than to rule out defensiveness. This is a result of Cashel et al.'s (1995) finding that some defensive respondents scored within normal limits on this index. A cutoff score of 70 was used for this study.

The categories of validity were used for interpretive purposes and also to separate out valid and invalid groups for statistical analysis. The "interpret with caution" group was included with the "invalid" group as the literature has found that the lower cutoff point results in a more accurate classification of invalid protocols.

Validity The convergent validity of the PAI has been investigated through concurrent administrations of major clinical instruments. In addition, diagnostic and clinical information from the treating clinician were compared with their hypothesised PAI correlates (Morey, 1996). These studies have found the PAI scales correlate strongly with most major tests in diagnoses. Some of the examples provided by Morey (1996) are presented below.

The Anxiety (ANX) scale correlated with a number of measures including the NEO-PI Neuroticism (r =.76), the STAI Trait Anxiety Inventory (r =.73) and the Wiggins Depression (r =.76). The Depression (DEP) scale correlated with the Beck Depression Inventory (r =.81), the Hamilton Depression Scale (r =.78), the MMPI D scale (I =.66), the NEO-PI Neuroticism (r =.69) and Depression (r =.70) scales and the Beck Hopelessness scale (r = .67). The Paranoia (PAR)

scale correlated with the MMPI Paranoid personality disorder scale (r = .70) and the Wiggins Psychoticism scale (r = .60).

The Borderline Features (BOR) scale correlated with the MMPI Borderline Personality scale (r = .77) and the NEO-PI Neuroticism scale (r = .67). The PAI Antisocial Features (ANT) scale correlated with the Hare Psychopathy Scale (r = .82) and the MMPI Antisocial personality disorder scale (r = .77).

7.6.1 NEO-PI-R

The NEO-PI-R is administered to all applicants to the New Zealand Police and is the primary personality test at present alongside a cognitive abilities test.. The NEO-PI-R consists of 240 items forming five domain scales and 30 facet scales and is intended to be a measure of normal personality using the Five Factor Model.

Reliability Coefficient alphas for the domain scales range from .86 for Agreeableness to .92 for Neuroticism. Alpha figures for the facet scales are lower with all but three over .6 and most over .7 (Costa & McCrae, 1992a).

Validity The NEO-PI-R does not contain any validity scales to assess for response sets, but does have three questions that ask if respondents have answered honestly and accurately, responded to all items and entered responses in the right place (Costa & McCrae, 1992a).

The NEO-PI-R has been validated against a range of other instruments including the PAI. Costa and McCrae (1992b) found a number of correlations between NEO-PI-R domain scales and PAI scales. For example, the NEO-PI-R Neuroticism scale correlated with Anxiety (r = .63), Depression (r = .40), Paranoia (r = .43), Schizophrenia (r = .51), and Borderline Features (r = .67). Correlations with the PAI were found for the other four domain scales also.

7.6.2 RNZPC Performance Data

The performance criterion used was the final grade achieved by recruits at the completion of their training. A variety of assessments contributed to this grade including short written tests and longer exams, which tested knowledge of police procedures and relevant laws, and practical tests of skills such as driving and the use of firearms.

Chapter 8: Results

8.0 Overview

The data presented in this chapter are intended to provide information relevant to the evaluation of the PAI as a selection instrument for the New Zealand Police. The first section deals with the issue of reliability, while the second is concerned with the various aspects contributing to test validity. The third section of this chapter evaluates the relationship between demographic factors and test scores in order to check for possible biases against any group. The fourth section looks at the scores of the recruits as a group and relates this to the normative community sample. The last sections deal with an assessment of safety issues and a brief description of elevated PAI profiles found in this sample.

The amount of data generated by the analyses of two multi-scale tests and performance measures meant that it was not practical to present all data in this chapter. Appendix A contains tables of correlations and other data that has been omitted here.

The PAI and NEO-PI-R contain a number of full and subscales, and have been referred to in this chapter using acronyms and full titles. A description of the content of each scale and what it is intended to measure is contained in Appendix C.

8.1 Reliability

The first objective of this study was to evaluate the reliability of the PAI. The internal consistency of the test was calculated using the coefficient alpha measure and inter-item correlations. These results are presented in Tables 4 and 5.

The data from the present study was compared with data from Morey's (1991) normative study. He used a census-stratified sample of the North American community with an *N* of 1000, as well as a sample of African Americans and college students. The community sample was used for comparative purposes, as recommended by Morey (1991), because it represents the broadest range of people.

Table 4 shows the internal consistency of the PAI full scales for the police recruit sample (*N* = 127), as compared to the normative community sample (*N* = 1000). The median alpha for the police recruit sample across all scales was .70 compared to .81 for the community sample. Alpha coefficients are lower for the police sample across all scales with DRG (Drug Problems) and SUI (Suicidal Ideation) being particularly reduced. All the validity scales except PIM (Positive Impression Management) are also considerably lower for the police sample. All but two clinical scales have demonstrated an adequate level of reliability with alphas remaining above .7, which is considered to be a minimum requirement (Kline, 1993). However, alphas of .85 are considered to be ideal for a measure for personnel selection (Gatewood & Field, 1998), and no scale has achieved this.

Mean inter-item correlation for each scale reveals to some extent the degree to which items are correlated without the effect of scale length. This correlation can demonstrate the independence of item content, showing the degree to which items measure differing aspects of a construct. Broad, heterogenous constructs are likely to have low inter-item correlation due to the wide range of item content needed to measure differing aspects. Correlations between items

are lower for the police recruit sample (median of .16) than for the community sample (median of .22), and only four of the scales exceeded .2 on this measure. This indicates that the recruits' responses to items within scales were more varied than for the normative sample. The very low level of responding to items on scales such as SUI and DRG will have contributed to this, as the endorsement of a few items on these scales while the rest scored zero will have produced low inter-item correlation.

Table 5 shows that for most subscales, alpha coefficients were lower for the police sample, with subscales such as SOM-H (Health Concerns) and ANX-P (Physiological Anxiety) being particularly reduced. There were some subscales such as ANT-S (Stimulus Seeking) and MAN-G (Grandiosity), which had higher alphas although not by a large margin. The median alpha for the police recruit sample was .63 compared to .73 for the community sample. The subscales were expected to have lower alphas than the full scales partly as a function of the length of the scales (Kline, 1993).

Inter-item correlations were higher than for full scales, as expected, because the subscale items are a more specific measure of particular domains (Morey, 1996). The correlations were generally lower for the police sample (median of .19) compared to the community sample (median of .27), indicating that the subscales were providing a less consistent measure of constructs for this population. There were some subscales, however, where correlations were higher, such as ANT-S (Stimulus Seeking), suggesting that the ANT-S items were responded to by the police sample in a more consistent manner.

Table 4. *Alpha Coefficients and Mean Inter-item Correlation for PAI Full Scales for the Police Recruit (N = 127) and Community Samples (N = 1000).*

	Coefficient Alpha		Mean Inter-item Correlation	
Scale	Police Recruit	Community	Police Recruit	Community
Validity Scales				
Inconsistency (ICN)	.35	.45	.04	.08
Infrequency (INF)	.40	.52	.09	.14
Negative Impression (NIM)	.46	.72	.14	.24
Positive Impression (PIM)	.73	.71	.24	.17
Clinical Scales				
Somatic Complaints (SOM)	.69	.89	.12	.26
Anxiety (ANX)	.84	.90	.18	.17
Anxiety-Related Disorders (ARD)	.70	.76	.11	.13
Depression (DEP)	.73	.87	.10	.24
Mania (MAN)	.81	.82	.16	.17
Paranoia (PAR)	.75	.85	.13	.20
Schizophrenia (SCZ)	.71	.81	.10	.17
Borderline Features (BOR)	.81	.87	.16	.22
Antisocial Features (ANT)	.82	.84	.16	.20
Alcohol Problems (ALC)	.70	.84	.21	.36
Drug Problems (DRG)	.32	.74	.07	.28
Treatment Scales				
Aggression (AGG)	.77	.85	.18	.19
Suicidal Ideation (SUI)	.49	.85	.20	.41
Stress (STR)	.54	.76	.14	.30
Nonsupport (NON)	.68	.72	.21	.25
Treatment Rejection (RXR)	.61	.76	.17	.28
Interpersonal Scales				
Dominance (DOM)	.70	.78	.16	.22
Warmth (WRM)	.71	.79	.17	.24
Median Value	.70	.81	.16	.22

Table 5. *Alpha Coefficients and Mean Inter-item Correlation for PAI Subscales for the Police Recruit (N = 127) and Community Samples (N = 1000).*

	Coefficient Alpha		Mean Inter-item Correlation	
Scale	Police Recruits	Community	Police Recruits	Community
Somatic Complaints (SOM)				
Conversion (SOM-C)	.75	.74	.27	.27
Somatization (SOM-S)	.41	.68	.09	.22
Health Concerns (SOM-H)	.29	.81	.12	.36
Anxiety				
Cognitive (ANX-C)	.73	.81	.23	.35
Affective (ANX-A)	.70	.73	.24	.27
Physiological (ANX-P)	.44	.74	.10	.29
Anxiety-Related Disorders (ARD)				
Obsessive-Compulsive (ARD-O)	.52	.56	.12	.14
Phobias (ARD-P)	.44	.58	.10	.15
Traumatic Stress (ARD-T)	.73	.81	.30	.35
Depression (DEP)				
Cognitive (DEP-C)	.65	.74	.20	.28
Affective (DEP-A)	.51	.80	.11	.36
Physiological (DEP-P)	.63	.71	.14	.23
Mania (MAN)				
Activity Level (MAN-A)	.48	.51	.11	.12
Grandiosity (MAN-G)	.76	.73	.30	.26
Irritability (MAN-I)	.78	.78	.32	.31
Paranoia (PAR)				
Hypervigilance (PAR-H)	.60	.64	.17	.19
Persecution (PAR-P)	.53	.76	.17	.30
Resentment (PAR-R)	.52	.66	.13	.21
Schizophrenia (SCZ)				
Psychotic Experiences (SCZ-P)	.40	.56	.09	.16
Social Detachment (SCZ-S)	.63	.79	.16	.33
Thought Disorder (SCZ-T)	.64	.73	.21	.27
Borderline Features (BOR)				
Affective Instability (BOR-A)	.63	.71	.21	.29
Identity Problems (BOR-I)	.56	.70	.19	.30
Negative Relationships (BOR-N)	.65	.63	.24	.22
Self-Harm (BOR-S)	.59	.62	.18	.22
Antisocial Features (ANT)				
Antisocial Behaviours (ANT-A)	.67	.73	.21	.27
Egocentricity (ANT-E)	.52	.63	.11	.20
Stimulus-Seeking (ANT-S)	.78	.69	.30	.23
Aggression (AGG)				
Aggressive Attitude (AGG-A)	.72	.74	.30	.32
Verbal Aggression (AGG-V)	.65	.67	.24	.25
Physical Aggression (AGG-P)	.52	.71	.22	.34
Median Value	.63	.73	.19	.27

8.2 Validity of PAI test protocols

Protocol validity refers to the extent to which the scores of respondents are influenced by response sets such as defensiveness or a random completion of the test. This issue was considered important when considering the implications of test scores, and was assessed using the validity scales incorporated in the PAI. The recruits were separated into two groups based on the likelihood of defensiveness or other response sets having influenced responding, and the mean *T* scores for the groups were compared. These results are presented in Tables 6 and 7.

Table 6. *Validity of PAI Protocols for the Police Recruit Sample (N = 127).*

Status	Frequency	Percent
Invalid – random	4	3.1
Invalid – defensive	25	19.7
Interpret with caution – defensive	45	35.4
Interpret with caution – defensive and random	13	10.2
Interpret with caution – random	5	3.9
Valid	35	27.6
Total	127	100.0

Table 6 shows the proportion of test protocols found to be valid according to the criteria set out in the methodology section of the present thesis. Only 27.6 % of test protocols were assessed as being valid with some confidence, with the remainder being possibly influenced by defensive or careless responding by participants. If Morey's (1996) study on the ability of PIM (Positive Impression Management) to correctly identify defensive protocols is used, there are still approximately 18% of the valid protocols misclassified for defensiveness (invalid classed as valid). The other protocols, which have been classed as invalid or to be interpreted with caution, may contain 30% that are misclassified (valid classed as invalid).

The results suggest that a considerable percentage of the police sample were attempting to present an overly favourable impression of their psychological wellbeing. However, the criteria used to judge protocol validity has been developed in other populations including college students and prison inmates. The efficacy of these measures and the appropriateness of cutoff scores used have not been demonstrated in a police population.

Table 7. *Significant Differences Between Mean PAI Scale Scores for Valid and Invalid Protocols for the Police Recruit Sample (N = 127).*

	Valid (*N*=35)		Invalid/Interpret with caution (*N*=92)		
Scale	Mean	*SD*	Mean	*SD*	Sig
Validity Scales					
Inconsistency (ICN)	47.63	6.72	46.88	8.99	
Infrequency (INF)	49.37	5.47	53.32	8.75	*
Negative Impression (NIM)	47.71	6.18	46.12	3.62	
Positive Impression (PIM)	48.23	5.41	60.00	6.08	**
Clinical Scales					
Somatic Complaints (SOM)	44.69	4.71	42.86	3.16	**
Anxiety (ANX)	49.94	6.09	45.05	5.89	**
Anxiety-Related Disorders (ARD)	48.00	6.32	42.60	6.29	**
Depression (DEP)	45.49	5.88	43.73	5.05	
Mania (MAN)	52.69	8.59	48.28	7.93	**
Paranoia (PAR)	49.34	6.37	45.89	6.04	**
Schizophrenia (SCZ)	48.83	5.08	45.12	6.50	**
Borderline Features (BOR)	50.80	7.17	44.30	5.79	**
Antisocial Features (ANT)	57.60	8.90	52.34	8.23	**
Alcohol Problems (ALC)	53.23	8.11	50.29	6.24	
Drug Problems (DRG)	47.89	4.92	47.96	6.29	
Treatment Scales					
Aggression (AGG)	48.74	7.53	45.37	5.67	**
Suicidal Ideation (SUI)	47.60	4.23	45.54	4.59	*
Stress (STR)	50.03	5.74	46.33	5.71	**
Nonsupport (NON)	46.51	7.04	43.76	7.40	
Treatment Rejection (RXR)	53.06	5.48	58.40	6.46	**
Interpersonal Scales					
Dominance (DOM)	49.37	7.80	50.90	6.51	
Warmth (WRM)	50.54	6.47	53.60	7.39	*

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 7 displays the mean *T* scores of valid protocols compared with invalid protocols (including those classified "interpret with caution"). Significant differences ($p < .05$) were found for the INF, PIM, SOM, ANX, ARD, MAN, PAR, SCZ, BOR, ANT, AGG, SUI, STR and WRM scales. Valid protocols scored higher on both clinical and treatment scales with the exception of RXR (Treatment Rejection). This scale is commonly elevated in a defensive profile, possibly indicating a reluctance to admit to a need for change (Morey, 1996).

8.3 Construct Validity

One objective of this study was to evaluate the construct validity of the PAI. This was done by correlating the PAI scales with each other and with the NEO-PI-R. In order to demonstrate construct validity, the correlations found should present in the following manner. The PAI scales should correlate in a positive direction with other measures of the same construct to show convergent validity. They should have a negative or zero correlation with scales measuring theoretically incompatible constructs in order to show discriminant validity (Kline, 1993). Because of the large number of correlations generated by this analysis, only a sample of these has been described in the text. All correlations produced may be found in Appendix A. The strength of correlations was classed as being high (.6 -.8), moderate (.4 -.6) or low (.2 -.4).

8.3.0 Interscale Correlation

PAI Interscale Correlations PIM (Positive Impression Management) was negatively correlated with all clinical scales and positively related to RXR (Treatment Rejection). This concurs with the finding that defensiveness is a negative influence on openness to treatment (Morey, 1996), and that PIM elevations may have a suppressive effect on responding (on clinical scales).

For clinical and treatment scales, all significant correlations were positive, indicating that increased scores on one scale were associated with increased scores on others. This was expected, as all of these scales measure pathological characteristics. The exception to this pattern is the RXR (Treatment Rejection) scale, which had a negative relationship with other scales, suggesting that those reporting less psychological difficulties were more likely to report being closed to the possibility of personal change being necessary.

ANX (Anxiety) had moderate to strong correlations with other scales with a significant anxiety component such as SOM (Somatic Complaints), ARD

(Anxiety Related Disorders), DEP (Depression), PAR (Paranoia), SCZ (Schizophrenia), and BOR (Borderline Features).

Other neuroticism related constructs (Somatic Complaints, Anxiety-Related Disorders and Depression) intercorrelated positively at a low to moderate level. Increased levels on one scale were associated with increased scores on the other scales.

Scales measuring psychosis-related constructs (Mania, Paranoia, Schizophrenia) intercorrelated at a low level, considerably lower than for the community sample. This may be related to the restricted range of responding on these scales. Moderate to strong correlations were found between the BOR (Borderline Features) scale and scales measuring psychoticism and neuroticism. This reflects the disorder's presence on the boundary between these two facets of personality.

PAI and NEO-PI-R Interscale Correlations The PAI PIM (Positive Impression Management) scale had a low to moderate negative relationship with NEO-PI-R Neuroticism scales, and a low to moderate positive association with Conscientiousness scales. This indicated that those scoring high on PIM when taking the PAI tended to score lower on NEO-PI-R scales measuring undesirable characteristics such as Neuroticism and higher on desirable characteristics such as Conscientiousness. The PAI scale RXR (Treatment Rejection) had a moderate negative association with Neuroticism scores on the NEO-PI-R. This indicated that increased scores on neuroticism related scales were associated with an increased willingness to admit to a need for personal change.

The PAI neurotic scales correlated at a low to moderate level with NEO-PI-R Neuroticism facet scales. ANX (Anxiety) correlated with the N1 (Anxiety) scale at a moderate level, but correlated equally strongly with N3 (Depression). This correlation was stronger than the relationship between DEP (Depression) and N3. ARD (Anxiety Related Disorders) also had a stronger relationship to N3 than did DEP. Although the PAI scales mentioned did correlate with

theoretically concordant NEO-PI-R scales, the relative strength of these correlations was not expected. It was expected that scales on each test measuring the same construct would correlate higher with each other than for other scales, and demonstrated a possible lack of discriminant validity for these scales.

Other PAI clinical scales were associated with increased NEO-PI-R Neuroticism scores. BOR (Borderline Features) in particular had low to moderate correlations with all N scales.

Among the PAI treatment scales, AGG (Aggression) was found to have a moderate correlation with the N2 (Angry Hostility) scale. Notably, DRG (Drug Problems) was not correlated significantly with any NEO-PI-R scale. This may have been due to the low level of endorsement of items on this scale. RXR (Treatment Rejection) had a moderate negative relationship to the Neuroticism scales and a low positive association with Conscientiousness scales.

The PAI's interpersonal scales of DOM (Dominance) and WRM (Warmth) both had moderate positive associations with similar NEO-PI-R scales. DOM correlated with E3 (Assertiveness) and E4 (Activity) and also negatively correlated with A5 (Modesty). WRM was particularly correlated with E1 (Warmth) and A3 (Altruism).

The correlations found between PAI and NEO-PI-R scores were generally lower than those reported by Morey (1991). In particular, many of the correlations between PAI full scales and the NEO-PI-R domain scale of Neuroticism were considerably lower than Morey's (1991) data. On scales such as MAN (Mania), PAR (Paranoia) and ANT (Antisocial Features), however, correlations with Neuroticism found in the present study were similar to, or exceeded those, in a study by Costa and McCrae (1992). PAI scales such as DEP (Depression) and ANX (Anxiety) showed some lack of discriminant validity when compared to the NEO-PI-R, as they correlated higher with scales measuring a different construct.

8.3.1 Structure of the PAI

The structure of the PAI was examined using principal components extraction followed by varimax rotation, and performed on the correlation matrix generated by the police recruit sample. This method was used to replicate the principal components analysis performed by Morey (1991) in order to make some descriptive comparisons with his findings.

There has been some debate in the literature (Boyle and Lennon, 1994; Morey, 1995;Conger and Conger, 1996) as to the appropriateness of this methodology. In particular, the use of an orthogonal rotation rather than an oblique method has been questioned by Boyle and Lennon (1994) due to the size of intercorrelations found in PAI factors produced by Morey (1991). As Morey (1996) mentions, the debate over the "best" method of extraction and rotation of factors has been ongoing for some time without clear resolution. Tabachnick and Fidell (1989) comment that despite the large number of techniques available, the practical differences between them are often low. They describe principal components extraction with varimax rotation as being the most common initial method of analysis.

Before applying factor analytic techniques to the police sample, some assumptions were checked to ensure the suitability of this analysis. A minimum of five cases per variable is preferred for factor analysis (Tabachnick & Fidell, 1989). After removal of univariate outliers (those cases with scores more than three standard deviations from the mean of the sample), *N* was 124 giving 5.6 cases per variable with all 22 PAI scales included in the analysis.

A principal components analysis was then conducted to check for variables that obtained a measure of sampling accuracy lower than or equal to .50. On this basis ICN (Inconsistency) and DRG (Drug Problems) were removed from the analysis.

For the remaining scales, factorability of the matrix was assessed using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin Measure of Sampling

Adequacy of the matrix. If the figure produced by Bartlett's test is large and significant and the KMO result exceeds .6, factorability is assumed. In this case, both tests produced a satisfactory result.

The correlation matrix was also checked for outlier variables that had low correlations with all other variables, and the INF (Infrequency) scale was excluded on this basis.

Principal Components Analysis was used as the extraction technique. Components with eigenvalues greater than one were extracted which resulted in a four factor solution. The fourth factor extracted had an eigenvalue of 1.02 making it marginal for inclusion.

Only variables with loadings of .55 or greater were interpreted. Tabachnik and Fidell (1989) state that the choice of a cutoff point for loadings is based on researcher preference and that often a loading point is chosen to make the solution easier to interpret. This was applicable in the present study as choosing .55 allowed the four factors to become clearer with less overlap of variable loadings. Comrey (1973, cited in Tabachnik & Fidell, 1989) rates loadings of .55 or greater as good with .45 being fair and .63 very good.

Table 8. *Orthogonal PAI Factor Pattern for the Police Recruit Sample (N = 124).*

Scale	Factor 1	Factor 2	Factor 3	Factor 4
ANX (Anxiety)	.83			
RXR (Treatment Rejection)	-.72			
DEP (Depression)	.70			
ARD (Anxiety-Related Disorders)	.60			
PIM (Positive Impression Management)	-.56			
BOR (Borderline Features)	.56			
MAN (Mania)		.80		
ANT(Antisocial Features)		.72		
DOM (Dominance)		.67		
AGG (Aggression)		.59		
ALC (Alcohol Problems)				
NIM (Negative Impression Management)				
WRM (Warmth)			-.74	
NON (Nonsupport)			.61	
PAR (Paranoia)			.60	
SCZ (Schizophrenia)			.56	
SOM (Somatic Complaints)				.74
SUI (Suicidal Ideation)				.63
STR (Stress)				
% variance	34.20	12.24	7.31	5.36
Eigenvalue	6.50	2.33	1.39	1.02

Factor 1 is made up of negative loadings on Positive Impression Management and Treatment Rejection and positive loadings on Anxiety, Depression, Anxiety Related Disorders and Borderline Features. This factor could be described as representing Neuroticism with the three of the four neuroticism related scales included. A common factor seems to be the experience of psychological distress. The negative loadings on PIM and RXR suggest that this is associated with a non-defensive response style, and openness to the need for change. This factor is similar to Morey's (1991) Factor 1, which was also characterised by psychological distress and affective disruption.

Factor 2 consists of positive loadings on Mania, Antisocial Features, Dominance and Aggression. This factor seems to represent interpersonal characteristics that involve aggression, dominance and a disregard for others. Mania may be included in this factor through the contributions of MAN-I (Irritability) and MAN-A (Activity Level) which would seem to fit with this pattern. Morey's (1991) Factor 3 is similar to this with the exception of Antisocial Features.

Factor 3 is defined by a negative loading on Warmth and positive loadings on Non-support, Paranoia and Schizophrenia. This is similar to Factor 4 found by

Morey (1991, p. 137) and was described by him as involving "social detachment and a touchiness and sensitivity in social relationships".

Factor 4 does not seem to be easily interpretable with the scales measuring Somatic Complaints and Suicidal Ideation loading strongly on this scale. Both these scales had a highly skewed distribution with many recruits scoring zero. The infrequency of higher scores on these scales may be related to some common element of idiosyncratic responding to the test.

8.4 Criterion-Related Predictive Validity

This aspect of analysis evaluates the ability of the PAI and NEO-PI-R to predict future performance. In this study, the final grade achieved during training at the Royal New Zealand Police College was used as the criterion. Because of the apparent impact of protocol validity on PAI scores, the effect of response sets on predictive validity was investigated.

8.4.0 Effects of Response Sets on Predictive Validity

Response sets and defensiveness in particular have been thought to have little effect on the predictive validity of tests used in personnel selection (Hough et al., 1990). This was partially investigated by looking at the univariate correlations between PAI and NEO-PI-R scores and performance for groups classed as valid or invalid protocols.

Table 9. *Correlations Between Final Grade Obtained at the RNZPC and Scores on the PAI and NEO-PI-R for all Recruits and Valid and Invalid Protocols.*

Scales	Total	Invalid	Valid
PAI full scales	SOM -.23* ANX -.22* DEP -.25** SUI -.18* DOM .20* *N* = 126	ANX -.21* DEP -.30** DOM .29** *N* = 91	INF -.39* SUI -.34* *N* = 35
PAI subscales	SOM-S -.18* SOM-H -.23* ARD-P -.25** DEP-C -.21* DEP-P -.20 PAR-P -.22* BOR-S -.18* *N* = 126	ANX-C -.26* ARD-P -.23* DEP-C -.30** DEP-P -.26* PAR-P -.34** PAR-R -.24* *N* = 91	SOM-H -.36* BOR-S -.37* *N* = 35
NEO-PI-R facet scales	N5 -.26** O5 .22* C4 .20* *N* = 106	C4 .23* *N* = 76	N5 -.40* *N* = 30

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

KEY PAI Full Scales *INF* Infrequency *SOM* Somatic Complaints *ANX* Anxiety *DEP* Depression *SUI* Suicidal Ideation *DOM* Dominance **PAI Subscales** *SOM-S* Somatization *SOM-H* Health Concerns *ANX-C* Cognitive *ARD-P* Phobias DEP-C Cognitive DEP-P Physiological *PAR-P* Persecution *PAR-R* Resentment *BOR-S* Self-Harm **NEO-PI-R Facet Scales** N5 Impulsiveness O5 Ideas C4 Achievement Striving

Table 9 shows the significant correlations found between test scores and final grade achieved at the RNZPC for valid, invalid and all protocols. With all participants included in the analysis, and looking at the relationship between PAI full scale scores and performance, five scales correlated significantly. These results suggest that higher levels of somatic complaints, anxiety, depression, and suicidal ideation were significantly associated with lower performance and that dominance was associated with higher performance.

Anxiety, depression and dominance were significantly correlated with performance for the invalid protocols. However, the association between depression and dominance and the performance criterion strengthened compared to the group including all participants. Only suicidal ideation and infrequent responding correlated significantly for the valid group with higher scores on both scales being associated with lower performance.

When the correlation between PAI subscales and performance was examined, higher levels of conversion symptoms (SOM-C), health concerns (SOM-H), cognitive and affective dimensions of anxiety (ANX-C; ANX-A), phobias (ARD-

P), cognitive and physiological dimensions of depression (DEP-C; DEP-P), paranoia (PAR-P), and self-harm (BOR-S) were associated with lower levels of performance with all participants included.

With only the participants included whose protocols were classed as invalid or to be interpreted with caution, higher levels of cognitive anxiety (ANX-C), phobias (ARD-P), cognitive and physiological dimensions of depression (DEP-C; DEP-P), paranoia (PAR-P), and resentment (PAR-R) correlated with lower levels of performance. Higher levels of health concerns (SOM-H) and self-harm (BOR-S) correlated with lower performance for the valid group and were the strongest correlations found for PAI subscales.

Looking at the relationship between NEO-PI-R facet scales and performance, and with all participants included in the analysis, higher levels of impulsiveness (N5) was associated with lower performance, while ideas (O5), and achievement-striving (C4) was associated with higher performance.

With only the participants included whose protocols were classed as invalid or to be interpreted with caution, higher achievement-striving (C4) scores were significantly correlated with higher performance. With only valid protocols included, higher impulsiveness (N5) was associated with lower performance. This was the highest correlation found for NEO-PI-R facet scales. It is interesting to note that impulsiveness is a negative trait consistent with a group who are more likely to admit to negative psychological issues. For the invalid group, it was a positive psychological quality that predicted performance.

It appears that response sets may have some influence on predictive validity when univariate correlations are used. The valid group produced the fewest significant correlations, but those found were higher than for the invalid and combined group. All scales found to be significant for this group were negative psychological characteristics. The invalid group produced correlations that in most cases were of similar or greater magnitude than for the combined group. The correlations found included some positive psychological characteristics.

8.4.1 Prediction of RNZPC Final Grade with the PAI and NEO-PI-R

A series of hierarchical multiple regressions were carried out to examine the relationship between test scores and final performance at the RNZPC, while controlling for demographic variables (age, gender, ethnicity, marital status) and the validity of test protocols. Rather than present results for all scales, only those that had demonstrated a significant Pearson correlation with final performance were used in the regression analysis.

Before analysis, the variables were screened for assumptions of statistical analysis. The suggestion of Tabachnik and Fidell (1989) that conventional, but conservative alpha levels be used (eg. $p<.001$) to evaluate the significance of skewness and kurtosis was followed. Variables were converted to *Z* scores and those individuals with values over 3 or less than -3 were identified as univariate outliers and were deleted. This significantly improved the distributions. Transformations were applied to those variables for which elimination of outliers had not reduced skewness and kurtosis to a satisfactory level. Of these variables (SOM, SUI, SOM-S), only SOM could be improved, although the levels of skewness and kurtosis were still outside preferred parameters.

Tabachnick and Fidell (1989) suggest that there should be no less than 5 cases per variable and that more are needed if the effect sizes are small, there is substantial measurement error, or the DV is skewed. Cases per variable in this analysis ranged from 6 to 12. Analyses for PAI full and subscales, NEO-PI-R domain and facet scales, demographics and protocol validity (valid/invalid) are reported in Tables 10 to 17.

Table 10. *Hierarchical Multiple Regression of Demographic, Protocol Validity and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2 change for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 124).*

Predictor	Step 1	Step 2
Demographics:		
Age	-.21	-.21
Gender	.12	.18
Ethnicity	-.08	-.07
Marital Status	-.01	-.02
Protocol Validity	.05	-.09
PAI Full Scale Scores:		
SOM (Somatic Problems)		.01
ANX (Anxiety)		-.11
DEP (Depression)		-.09
SUI (Suicidal Ideation)		-.18
DOM (Dominance)		.20*
R	.26	.44**
R^2	.07	.19
Adjusted R^2	.03	.12
R^2 change	.07	.12**

*. $p < .05$, ** $p < .01$.

Analyses for PAI full scales are presented in Table 10. The *R* for regression (with final grade as DV) was not significantly different from zero after step 1. This indicates that demographics and protocol validity were not significant predictors of final grade on their own. With the addition of the PAI scales scores to the equation in step 2, $R^2 = .19$, $F(5, 110) = 2.6$, $p < .01$. The DOM scale contributed significantly to prediction of final grade with a beta of .20 ($p < .05$). The other scales that were significantly associated with final grade in univariate analysis have become non-significant here. This is a function of the intercorrelation between PAI scales. Altogether, the PAI scores accounted for 12% of the variation in final grade.

Table 11. *Hierarchical Multiple Regression of Demographic, Protocol Validity and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 121).*

Predictor	Step 1	Step 2
Demographics:		
Age	.20*	.30**
Gender	.11	.17
Ethnicity	-.09	-.06
Marital Status	-.02	-.02
Protocol Validity	.08	-.09
PAI Subscale Scores:		
SOM-S (Somatization)		-.10
SOM-H (Health Problems)		-.02
ANX-C (Cognitive)		-.06
ANX-A (Affective)		.06
ARD-P (Phobias)		-.14
DEP-C (Cognitive)		-.01
DEP-P (Physiological)		-.02
PAR-P (Persecution)		-.22*
BOR-S (Self-Harm)		-.19
R	.26	.47*
R^2	.07	.22
Adjusted R^2	.03	.12
R^2_{change}	.07	.15*

*. $p < .05$, ** $p < .01$.

Analyses for PAI subscales are presented in Table 11. Although age correlated significantly with final grade in both steps, the demographic and protocol validity variables together did not significantly predict final grade. The age variable also shows a suppression effect (Smith, Ager, and Williams, 1992), which makes interpretation somewhat more difficult. After step 2, with the addition of the PAI subscales to the equation, $R^2 = .22$, $F(9, 106) = 2.12$, $p < .05$. The subscale PAR-P was the only scale to contribute significantly to the prediction of final grade and 22% (12% adjusted) of variation on final grade was predicted by knowing scores on these variables.

Table 12. *Hierarchical Multiple Regression of Demographic, Protocol Validity and NEO-PI-R Domain Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, adjusted R^2, and R^2 change (N = 106).*

Predictor	Step 1	Step 2
Demographics:		
Age	-.21*	-.23*
Gender	.07	.10
Ethnicity	-.12	-.09
Marital Status	-.02	-.04
Protocol Validity	.11	.05
NEO-PI-R Domain Scale Scores:		
N (Neuroticism)		-.04
C (Conscientiousness)		.19
R	.27	.34
R^2	.08	.12
Adjusted R^2	.03	.05
R^2 change	.08	.04

*. $p < .05$, ** $p < .01$.

Analyses for NEO-PI-R domain scales are presented in Table 12. Although age correlated significantly with final grade in step 1, the demographic and validity variables together did not significantly predict final grade. After step 2, with the addition of the NEO-PI-R domain scales to the equation, $R^2 = .15$, $F(7, 98) = 1.83$, $p = .09$. The addition of these variables did not result in a significant increment in R^2. Neither domain scale contributed significantly to the prediction of final grade.

Table 13. *Hierarchical Multiple Regression of Demographic, Validity and NEO-PI-R Facet Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} (N = 106).*

Predictor	Step 1	Step 2
Demographics:		
Age	-.21*	-.17
Gender	.07	.07
Ethnicity	-.16	-.06
Marital Status	-.02	-.02
Protocol Validity	.11	.05
NEO-PI-R Facet Scale Scores:		
N5 (Impulsiveness)		-.18
O5 (Ideas)		-.11
C4 (Achievement Striving)		-.11
R	.27	.39*
R^2	.08	.15
Adjusted R^2	.03	.08
R^2_{change}	.08	.07*

*. $p < .05$, ** $p < .01$.

Analyses for NEO-PI-R facet scales are presented in Table 13. Although age correlated significantly with final grade in step 1, the demographic and validity variables together did not significantly predict final grade. The main effects of age appear to be partially mediated by the NEO-PI-R facet scale scores. After step 2, with the addition of the NEO-PI-R facet scales to the equation, $R^2 = .15$, $F(3, 97) = 2.11$, $p < .05$. The addition of these variables resulted in a significant increment in R^2 (R^2 change = .07, $p < .05$). None of the facet scales contributed significantly to the prediction of final grade. Altogether 15% (8% adjusted) of variation on final grade was predicted by knowing the score on these variables.

In the preceding two-step regressions, demographic and protocol validity variables were not significant predictors of final grade on their own. PAI full scales predicted 12% of the variation in final grade, as did the PAI subscales. NEO-PI-R domain scales did not significantly predict final grade while the facet scales predicted 8%of final grade variance. The following regressions are composed of three steps, and were intended to investigate possible incremental validity of the PAI over the NEO-PI-R.

Table 14. *Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Domain Scales, and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 102).*

Predictor	Step 1	Step 2	Step 3
Demographics:			
Age	-.22*	-.23*	-.23*
Gender	.11	.14	.20
Ethnicity	-.06	-.05	-.05
Marital Status	-.05	-.07	-.04
Protocol Validity	.07	.11	-.06
NEO-PI-R Domain Scale Scores:			
N (Neuroticism)		.22	.26
C (Conscientiousness)		.12	.14
PAI Full Scale Scores:			
SOM (Somatization)			-.06
ANX (Anxiety)			-.28
DEP (Depression)			.04
SUI (Suicidal Ideation)			-.17
DOM (Dominance)			.15
R	.26	.30	.47*
R^2	.07	.09	.22
Adjusted R^2	.02	.02	.12
R^2_{change}	.07	.03	.13*

*. $p < .05$, ** $p < .01$.

The relative influence of NEO-PI-R domain scales and PAI full scales on final grade is presented in Table 14. This was analysed using demographic and validity variables in step 1, NEO-PI-R scales in step 2, and PAI full scales in step 3. The first step showed age to be the only demographic variable to have a significant influence on final grade scores, although the overall regression was not significant at this point. The R for regression was not significantly different from zero after step 2 with the addition of the NEO-PI-R scores ($R^2 = .09$, $F(7, 94) = 1.37$, $p = .23$). Using the PAI full scales in the final step resulted in a R value significantly different from zero, $R^2 = .22$, $F(12, 89) = 2.12$, $p < .05$. The addition of the PAI scales to the equation resulted in a significant increment in R^2 ($R^2_{change} = .13$, $p < .05$). None of the individual PAI scales correlated significantly with final grade on their own. Together the variables accounted for 23% (12% adjusted) of the variability in final grade scores.

Table 15. *Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Facet Scale, and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 101).*

Predictor	Step 1	Step 2	Step 3
Demographics:			
Age	-.15	-.14	.24*
Gender	.11	.10	.13
Ethnicity	-.05	-.02	-.02
Marital Status	-.06	-.04	.05
Protocol Validity	.10	.06	-.08
NEO-PI-R Facet Scale Scores:			
N5 (Impulsiveness)		-.17	-.07
O5 (Ideas)		.09	.16
C4 (Achievement Striving)		.11	-.08
PAI Subscale Scores:			
SOM-S (Somatization)			-.03
SOM-H (Health Concerns)			-.12
ANX-C (Cognitive)			-.07
ANX-A (Affective)			-.02
ARD-P (Phobias)			-.10
DEP-C (Cognitive)			.03
DEP-P (Physiological)			-.02
PAR-P (Persecution)			-.23*
BOR-S (Self-Harm)			.13
R	.21	.34	.51
R^2	.05	.11	.26
Adjusted R^2	-.01	.04	.11
R^2_{change}	.05	.07	.15

*. $p < .05$, ** $p < .01$.

The relative influence of NEO-PI-R facet scales and PAI subscales on final grade is presented in Table 15. This was analysed using demographic and validity variables in step 1, NEO-PI-R scales in step 2 and PAI subscales in step 3. The first step showed demographic and validity variables having no significant influence on final grade scores. The R for regression was not significantly different from zero with the addition of the NEO-PI-R scores (R^2 = .11, $F(3, 92) = 1.45$, $p = .19$). Using the PAI subscales in the final step resulted in a R value that was not significantly different from zero, $R^2 = .26$, $F(9, 83) = 1.73$, $p = .053$. The addition of the PAI scales to the equation resulted in a non-significant increment in R^2 ($R^2_{change} = .15$, $p = .053$). Together the variables accounted for 23% (12% adjusted) of the variability in final grade scores.

Table 16. *Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Facet Scale, and PAI Full Scale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2_{change} for Respondents Scoring Within Three Standard Deviations on PAI Full Scales (N = 102).*

Predictor	Step 1	Step 2	Step 3
Demographics:			
Age	-.16	-.14	.24*
Gender	.12	.11	.13
Ethnicity	-.05	-.03	-.05
Marital Status	-.04	-.03	.01
Protocol Validity	.07	.02	-.14
NEO-PI-R Facet Scale Scores:			
N5 (Impulsiveness)		-.12	-.12
O5 (Ideas)		.11	-.14
C4 (Achievement Striving)		.13	.05
PAI Full Scale Scores:			
SOM (Somatization)			-.06
ANX (Anxiety)			-.23
DEP (Depression)			-.02
SUI (Suicidal Ideation)			-.18
DOM (Dominance)			.14
R	.22	.33	.48*
R^2	.05	.11	.23
Adjusted R^2	.00	.03	.12
R^2_{change}	.05	.06	.13*

*. $p < .05$, ** $p < .01$.

The relative influence of NEO-PI-R facet scales and PAI full scales on final grade is presented in Table 16. This was analysed using demographic and validity variables in step 1, NEO-PI-R scales in step 2, and PAI full scales in step 3. The first step showed demographic and validity variables having no significant association with final grade scores. The R for regression was not significantly different from zero with the addition of the NEO-PI-R scores (R^2 = .11, $F(3, 93) = 1.38$, $p = .21$). Using the PAI full scales in the final step resulted in an R value significantly different from zero, $R^2 = .23$, $F(5, 88) = 2.05$, $p < .05$. The addition of the PAI scales to the equation resulted in a significant increment in R^2 (R^2 change = .13, $p < .05$). None of the individual PAI scales correlated significantly with final grade on their own. Together the variables accounted for 23% (12% adjusted) of the variability in final grade scores.

Table 17. *Hierarchical Multiple Regression of Demographic, Protocol Validity, NEO-PI-R Domain Scales, and PAI Subscale Scores on Final Grade at the RNZPC for Police Recruits Showing Standardised Regression Coefficients, R, R^2, Adjusted R^2, and R^2 change for Respondents Scoring Within Three Standard Deviations on PAI Subscales (N = 101).*

Predictor	Step 1	Step 2	Step 3
Demographics:			
Age	-.15	-.17	.26*
Gender	.11	.13	.18
Ethnicity	-.05	-.04	-.03
Marital Status	-.06	-.07	.07
Protocol Validity	.10	.06	-.08
NEO-PI-R Domain Scale Scores:			
N (Neuroticism)		-.03	.14
C (Conscientiousness)		.22	.20
PAI Subscale Scores:			
SOM-S (Somatization)			-.08
SOM-H (Health Concerns)			-.11
ANX-C (Cognitive)			-.11
ANX-A (Affective)			.00
ARD-P (Phobias)			-.13
DEP-C (Cognitive)			.03
DEP-P (Physiological)			-.01
PAR-P (Persecution)			-.23
BOR-S (Self-Harm)			.12
R	.21	.30	.49
R^2	.05	.09	.24
Adjusted R^2	-.01	.02	.09
R^2 change	.05	.04	.15

*. $p < .05$, ** $p < .01$.

The relative influence of NEO-PI-R domain scales and PAI subscales on final grade is presented in Table 17. This was analysed using demographic and validity variables in step 1, NEO-PI-R domain scales in step 2 and PAI subscales in step 3. The third step showed age as the only demographic variable to have a significant influence on final grade scores. The R for regression was still not significantly different from zero with the addition of the NEO-PI-R scores ($R^2 = .09$, $F(5, 95) = 1.24$, $p = .29$). Using the PAI subscales in the final step resulted in a R value that was not significantly different from zero, $R^2 = .24$, $F(16, 84) = 1.62$, $p = .08$. The addition of the PAI scales to the equation resulted in a non-significant increment in R^2 (R^2 change = .15, $p = .07$).

When added as a third step in the regression equations, PAI full scales demonstrated significant incremental validity over NEO-PI-R domain scales and facet scales separately. However, significant increments in predictive validity were not found for PAI subscales over NEO-PI-R domain or facet scales. NEO-PI-R scales did not demonstrate significant predictive validity in any of these regressions.

8.5 Test Fairness

The issue of test bias was assessed by looking for differences in test scores and correlations between different demographic groups. While any differences in scores do not necessarily indicate the test is biased, it highlights the possibility of an issue in this area.

9.5.0 Univariate Correlations Between Demographic Variables and Test Scores

Age A number of correlations were found between age of respondents and scores on the measures used. Only significant correlations ($p<.05$) are presented here.

The PAI scales Mania (r = -.29), Borderline Features (r = -.23), Antisocial Features (r = -.22) and Aggression (r = -.18) correlated with age ($p<.05$). The PAI subscales MAN-A (Activity Level) (r = -.25), MAN-G (Grandiosity) (r = -.25), MAN-I (Irritability) (r = -.19), BOR-A (Affective Instability) (r = -.20), BOR-S (Self-harm) (r = -.22), ANT-E (Egocentricity) (r = -.25) and ANT-S (Stimulus Seeking) (r = -.20) correlated with age ($p<.05$). All significant correlations are negative, indicating that pathology revealed in PAI scales decreases with the increased age of the respondent.

No significant correlations were found between NEO-PI-R domain scales and age. The NEO-PI-R facet scale E5 (Excitement Seeking) (r = -.32) correlated with age ($p<.05$). The only significant correlation found is negative and concurs with the PAI scores for a similar scale, ANT-S, which measures a desire for stimulation and excitement. Desire for stimulation and excitement decreases with increasing age.

Age correlated with final performance at the RNZPC (r = -.25). This suggests that increasing age is associated with lower grade.

Gender Significant differences found using the PAI are reported in Tables 18 and 19.

Table 18. *Significant Differences on PAI Scales Between Male (N = 104) and Female (N = 23) Police Recruits.*

	Male		Female	
Scale	Mean *T* score	SD	Mean *T* score	SD
Infrequency (INF)	51.29	7.92	56.25	8.08
Anxiety (ANX)	45.68	5.68	49.50	7.93
Anxiety-Related Disorders (ARD)	43.50	6.22	46.58	8.27
Borderline Features (BOR)	45.43	6.45	48.96	7.73
Stress (STR)	46.68	5.45	50.21	7.10
Anxiety – Cognitive (ANX-C)	45.24	4.26	44.58	2.87
Anxiety – Affective (ANX-A)	45.60	6.81	49.42	9.34
Anxiety - Physiological (ANX-P)	47.15	5.13	50.63	6.82
Phobias (ARD-P)	43.49	5.96	47.33	6.57
Traumatic Experiences (ARD-T)	45.19	5.25	47.79	7.29
Depression – Cognitive (DEP-C)	45.53	5.34	48.54	9.09
Activity Level (MAN-A)	50.28	8.43	54.71	8.51
Affective Instability (BOR-A)	43.89	5.85	47.25	5.95
Physical Aggression (AGG-P)	47.18	6.22	44.17	2.93

The results show the female recruits scored higher on measures of anxiety, phobia, traumatic experiences, depression, activity level, and affective stability. Males scored higher on a measure of physical aggression.

Table 19. *Significant Differences on NEO-PI-R Scales Between Male (N = 104) and Female (N = 23) Police Recruits.*

	Male		Female	
Scale	Mean *T* score	SD	Mean *T* score	SD
Openness to Experience (O)				
Fantasy (O1)	46.28	8.07	51.68	6.26
Values (O6)	51.20	5.75	54.95	6.52
Trust (A1)	51.45	6.71	46.42	9.46
Order (C2)	52.41	7.81	48.31	8.60

The results show female recruits scored higher on scales measuring imagination and readiness to examine values, and lower on scales measuring tendency to trust others and organisation.

Significant differences were found between genders for some RNZPC performance measures, but not for the final grade.

Ethnicity Significant differences were found between ethnic groups are reported in Table 20.

Table 20. *Significant Differences in Test Scores Between Maori (N = 17) and Caucasian (N = 110) Police Recruits.*

	Maori		Caucasian	
Scale	Mean *T* score	SD	Mean *T* score	SD
PAI Scales				
Somatic Complaints (SOM)	46.88	6.38	42.82	2.80
Depression (DEP)	48.18	8.60	43.60	4.37
Warmth (WRM)	57.18	6.28	52.07	7.18
NEO-PI-R Scales				
Depression (N3)	50.87	9.10	46.63	7.21
Impulsiveness (N5)	51.87	13.40	46.18	8.89
Aesthetics (O2)	53.53	7.84	46.75	7.85
Tendermindness (A6)	55.27	7.75	51.03	7.63

Maori recruits scored higher on measures of somatic complaints, depression, interpersonal warmth, impulsiveness, aesthetic appreciation, and concern for others. No significant differences were found between ethnic groups for final RNZPC performance grade.

Marital Status Significant differences in test scores between partnered and unpartnered recruits are reported in Table 21.

Table 21. *Significant Differences in Test Scores Between Partnered (N = 62) and Unpartnered Police Recruits (N = 65).*

	Unpartnered		Partnered	
Scale	Mean *T* score	SD	Mean *T* score	SD
PAI Scales				
Dominance (DOM)	51.77	7.05	49.25	6.55
Negative Relationships (BOR-N)	47.66	8.74	44.63	7.25
NEO-PI-R Scales				
Extraversion (E)	56.68	5.38	54.47	3.05
Excitement-Seeking (E5)	59.45	7.06	55.53	5.98
Positive Emotions (E6)	57.45	8.12	54.05	7.53

Unpartnered recruits scored higher on measures of dominance, negative relationships, extraversion, excitement-seeking and positive emotions.

Summary Demographic variables, and age and gender in particular, have some association with the scores of some of both tests are associated with scores on some PAI scales. Ethnicity does not appear to have a major relationship with

either test with differences occurring on only a few scales. In all cases, the differences in test scores based on demographic variables were small.

8.5.0 The Four-Fifths Rule

The ethnicity and gender of those identified as having highly elevated PAI scores (of 70 *T* or greater) was checked. *T* scores of 70 or greater are considered clinically significant by Morey (1996). Using the assumption that persons scoring in this range might not be selected for general or special police duties, the four-fifths rule was applied. According to this criterion, a test is deemed to have adverse impact on a group if their selection rate is 80% or less than the group with the highest selection rate (United States Equal Employment Opportunity Commission, 1978, cited in Hogg & Wilson, 1996).

The selection rates are calculated by dividing the number of recruits in a particular demographic group (ie. females or Maori) screened out by the test by the total number of recruits in that group. This selection rate is then compared to the demographic group with the highest selection rate. To meet the four-fifths criterion, the lowest selection rate should be no less than 80% of the highest selection rate.

The selection rate for females was 96.3 % of that for males. The selection rate of Maori was 76.2 % of that for Caucasians. While the test has not produced any difficulties based on gender, the selection rate for Maori is well below that of Caucasians and below the four-fifths criterion. However, none of the Maori in the group with clinically significant scores had elevations on scales that were found to be significantly different between the two ethnic groups as a whole. This suggests that the recruits screened out using this criterion differ in some way from other Maori as well as the Caucasian recruits.

8.6 PAI Scores for the Police Recruit Sample

Given that the recruits have been screened extensively during their application process, it was expected that the scores of this sample would differ from that of the normative sample. This section looks at the *T* scores of the police recruit sample compared to the normative community sample (Morey, 1991), which has a mean of 50 and a standard deviation of 10. This comparison is presented in Tables 22 and 23, and in Figures 1 and 2.

Table 22. *PAI Full Scale T scores for the Police Recruit Sample (N = 127).*

Scale	Mean	*SD*	Range	*SEM*
Validity Scales				
Inconsistency (ICN)	47	8.41	34-73	6.78
Infrequency (INF)	52	8.16	40-75	6.32
Negative Impression (NIM)	47	4.50	44-70	3.31
Positive Impression (PIM)	57	7.91	29-75	4.11
Clinical Scales				
Somatic Complaints (SOM)	43	3.73	39-62	2.08
Anxiety (ANX)	46	6.32	34-72	2.53
Anxiety-Related Disorders (ARD)	44	6.73	30-66	3.69
Depression (DEP)	44	5.32	35-67	2.76
Mania (MAN)	50	8.32	34-76	3.63
Paranoia (PAR)	47	6.30	29-67	3.15
Schizophrenia (SCZ)	46	6.35	33-74	3.42
Borderline Features (BOR)	46	6.82	32-71	2.97
Antisocial Features (ANT)	54	8.71	37-85	3.70
Alcohol Problems (ALC)	51	6.90	41-82	3.78
Drug Problems (DRG)	48	5.93	42-66	4.89
Treatment Scales				
Aggression (AGG)	46	6.39	32-72	3.06
Suicidal Ideation (SUI)	46	4.57	43-72	3.26
Stress (STR)	47	5.93	39-71	4.02
Nonsupport (NON)	45	7.38	37-72	4.17
Treatment Rejection (RXR)	57	6.63	31-72	4.14
Interpersonal Scales				
Dominance (DOM)	50	6.89	27-72	3.77
Warmth (WRM)	53	7.26	31-70	3.91

The police sample differed on the validity scales with ICN and NIM being depressed and INF, and particularly PIM, elevated. The most important difference is on the PIM scale, which measures a reluctance to admit minor flaws. Elevated PIM scores indicate that the participants may have been trying to present an overly favourable impression of themselves. The high mean score

for PIM was reflected in the large number of protocols (N = 92) being classed as invalid or to be interpreted with caution.

On most of the clinical scales, the recruits scored considerably lower than the community sample with SOM (Somatic Complaints), ARD (Anxiety-Related Disorders), and DEP (Depression) being particularly low, and ANT (Antisocial Features) and ALC (Alcohol Problems) slightly elevated. The recruits in general are reporting low levels of neurotic symptoms such as depression or somatic complaints, but were more likely to endorse items related to antisocial features or alcohol problems than the community sample.

For the treatment scales, the recruits again were low against the community sample with the exception of RXR (Treatment Rejection). The elevated RXR score concurs with a PIM elevation, as it is expected to be increased with defensive responding (Morey, 1996). It may however reflect a tendency for normal psychologically healthy people to see no need for treatment (Costa & McCrae, 1992).

The standard deviations of the police sample were much lower than for the community sample across all scales, with DRG (Drug Problems) and SOM (Somatic Complaints) having very low variance. A few scales such as MAN (Mania) and ANT (Antisocial Features) had somewhat higher standard deviations, with these scales also having means closer to that of the community. The range of scores on almost all scales shows that, although there is low variance as a group, there are still outliers that have scored more than two standard deviations above the community mean. This suggests that in spite of selection processes having produced a restriction in variance, the sample contains people who vary considerably not only from their fellow recruits, but from the broad community.

The standard errors of measurement calculated were generally low with most below four T points. This means that for a scale with a standard error of measurement value of four, the individual's true score has a 68% chance of falling with four T points either side of the observed score.

Table 23. *PAI Subscale T Scores for the Police Recruit Sample (N = 127).*

Scale	Mean	*SD*	Range	*SEM*
Somatic Complaints (SOM)				
Conversion (SOM-C)	45	5.17	43-87	2.59
Somatization (SOM-S)	42	3.97	38-54	3.05
Health Concerns (SOM-H)	45	4.03	40-62	3.40
Anxiety (ANX)				
Cognitive (ANX-C)	46	6.62	36-71	3.44
Affective (ANX-A)	46	7.46	34-70	4.09
Physiological (ANX-P)	48	5.63	58-69	4.21
Anxiety-Related Disorders (ARD)				
Obsessive-Compulsive (ARD-O)	47	7.95	25-70	5.51
Phobias (ARD-P)	44	6.24	31-62	4.67
Traumatic Stress (ARD-T)	46	5.75	41-67	2.99
Depression (DEP)				
Cognitive (DEP-C)	46	6.29	37-78	3.72
Affective (DEP-A)	45	4.83	39-66	3.38
Physiological (DEP-P)	45	6.63	36-69	4.03
Mania (MAN)				
Activity Level (MAN-A)	51	8.59	32-76	6.19
Grandiosity (MAN-G)	51	8.94	33-79	4.38
Irritability (MAN-I)	47	7.90	31-78	3.71
Paranoia (PAR)				
Hypervigilance (PAR-H)	49	7.65	28-77	4.84
Persecution (PAR-P)	46	5.45	39-69	3.74
Resentment (PAR-R)	47	7.07	30-61	4.90
Schizophrenia (SCZ)				
Psychotic Experiences (SCZ-P)	48	7.10	36-70	5.50
Social Detachment (SCZ-S)	47	6.10	36-64	3.71
Thought Disorder (SCZ-T)	47	6.93	37-75	4.16
Borderline Features (BOR)				
Affective Instability (BOR-A)	45	5.99	36-69	3.64
Identity Problems (BOR-I)	47	6.99	36-71	4.64
Negative Relationships (BOR-N)	46	8.13	34-72	4.81
Self-Harm (BOR-S)	50	8.43	37-72	5.40
Antisocial Features (ANT)				
Antisocial Behaviours (ANT-A)	50	8.33	39-75	4.79
Egocentricity (ANT-E)	53	7.99	39-82	5.54
Stimulus-Seeking (ANT-S)	58	10.51	40-86	4.93
Aggression (AGG)				
Aggressive Attitude (AGG-A)	42	7.51	34-73	3.97
Verbal Aggression (AGG-V)	52	7.73	31-74	4.57
Physical Aggression (AGG-P)	47	5.85	42-69	4.05

The means and standard deviations in Table 23 show on a more detailed level where the police sample may differ from the community sample. Of particular interest are the elevated scores on ANT-E (Egocentricity) and ANT-S (Stimulus-Seeking), which have increased the ANT full scale. ANT-A, which measures antisocial behaviour, is close to the community mean. The elevation in ANT-S (Stimulus-Seeking) may be the result of an age effect as 18-29 year olds have been found to score 7 *T* points higher than the rest of the normative sample

(Morey, 1996). It may also be a function of people being attracted to an occupation that appears to provide considerable excitement. Other elevated subscales include MAN-A (High Activity Level), MAN-G (Grandiosity) and AGG-V (Verbal Aggression). These however are only slightly above the community mean of 50. The remainder are lower than the community sample, but within one standard deviation.

The standard deviations of the recruit sample shows reduced variance compared to the community sample on all subscales except ANT-S. Despite the low variance of the group, it can be seen from the range that there are still some outliers for most subscales that are two standard deviations above the community mean.

Standard errors of measurement were higher for the subscales compared to the full scales, as a result of lower reliability. However, most were less than five *T* points with the exception of ARD-O (Obsessive-Compulsive), MAN-A (Activity Level), SCZ-P (Psychotic Symptoms), BOR-S (Self-Harm), and ANT-E (Egocentricity). These scales in particular need to be treated with caution when making interpretations, as there is a fairly wide confidence interval surrounding observed scores.

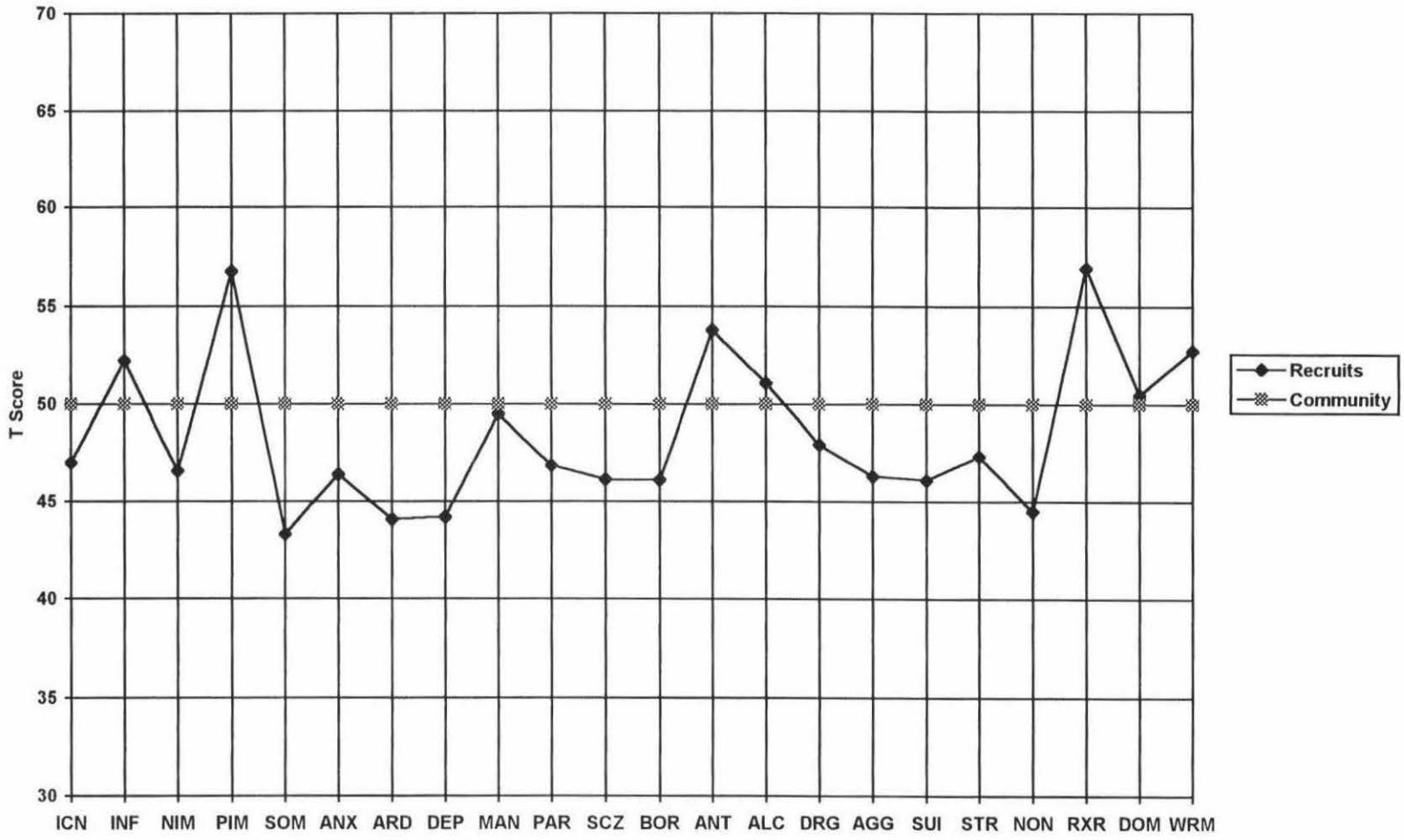

Figure 1. *Comparison of mean PAI full scale recruit profile (N =127) with normative community sample (N = 1000).*

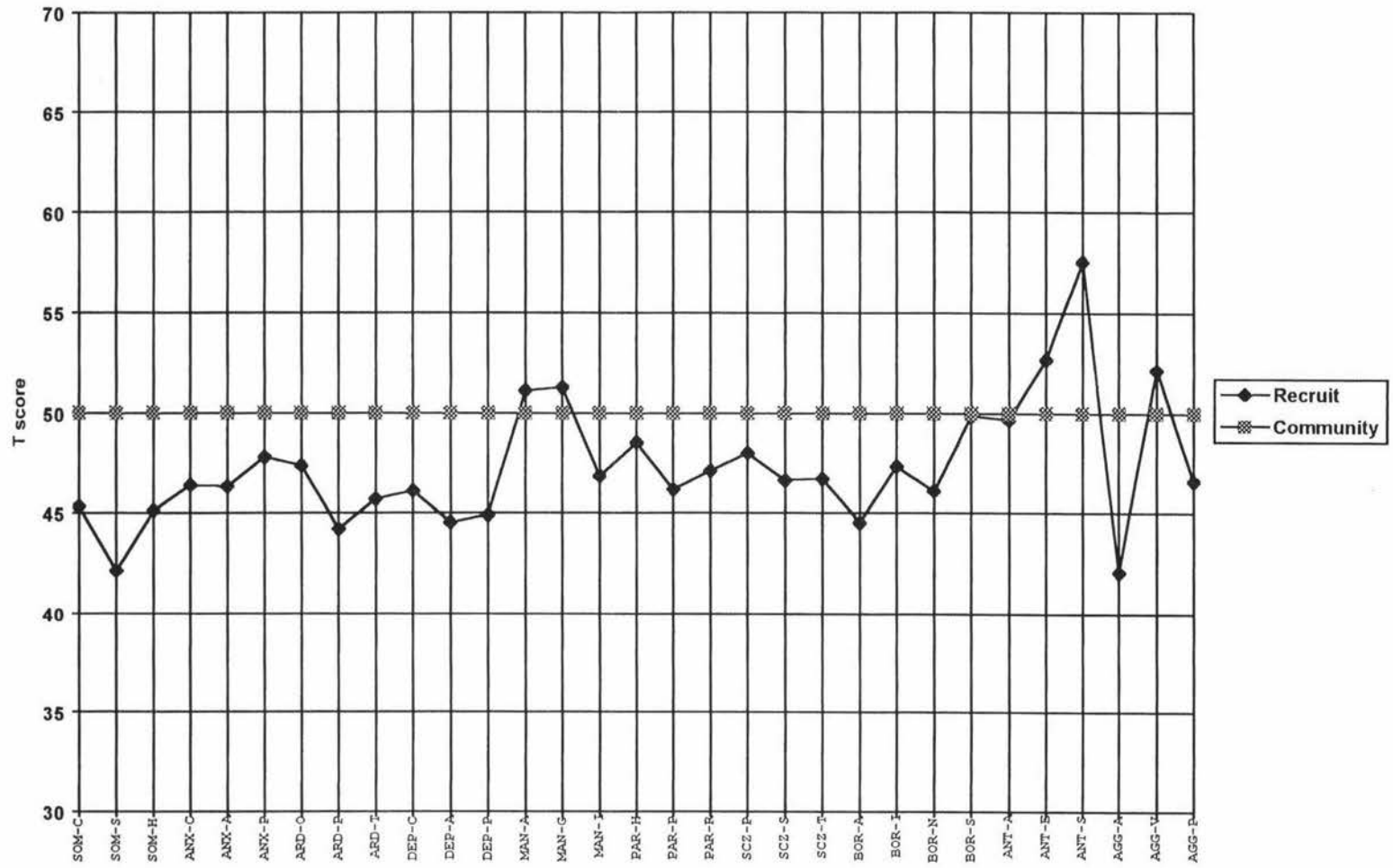

Figure 2. *Comparison of mean PAI recruit subscale profile (N = 127) with normative community sample (N =1000).*

8.7 Potential Danger to Self and Others

This aspect of analysis looks at what the PAI potentially reveals about the risk of recruits possibly harming themselves or others. Potential for aggression is measured using the AGG scales and the Violence Potential Index. Risk of self harm is assessed using the SUI scale and the Suicide Potential Index.

8.7.0 Potential for aggression

Table 24. *Elevations on PAI Scales and Indexes Related to Potential for Aggression for the Police Recruit Sample (N =127).*

Scale	No. scoring +1*SD*	No. scoring +2*SD*
Aggression (AGG)	5	1
Aggressive Attitude (AGG-A)	1	1
Verbal Aggression (AGG-V)	14	2
Physical Aggression (AGG-P)	5	0
Violence Potential Index	7	2

Table 24 shows the number of participants producing elevated scores on measures of aggression. Elevations have been categorised as exceeding one or two standard deviations from the mean of the community, which is the criterion, used for interpretation (Morey, 1996). The likelihood of inappropriate expression of anger rises above one standard deviation on these scales, with scores exceeding two standard deviations being of particular concern.

While there were a number of persons scoring above one standard deviation on aggression related scales, only two scored above two standard deviations. No recruit scored over two standard deviations on AGG-P, which is the most behavioural measure of aggression.

The Violent Potential Index is made up of 20 aspects of the PAI profile which fit with the literature available on the prediction of dangerousness and is intended to supplement the Aggression scales. Morey (1996) uses one and two standard deviation points above clinical means as thresholds to suggest moderate and marked risk for violent behaviour respectively. This corresponds to raw scores of nine and 17.

The recruit sample was low on this index with a mean of 1.05 raw points and a standard deviation of 1.44. The highest score obtained on this index was seven, which corresponds to a *T* score of 75 compared to community norms and 57 for clinical norms. Although this person would not be assessed as being a significant risk for violent behaviour, his/her marked deviation from community norms suggests that potential for aggression may be a concern. Samples of people with a history of violence such as assault convictions have obtained similar scores (Morey, 1996).

8.7.1 Potential for Self-Harm

The Suicide Potential Index is made up of 20 aspects of the PAI profile that are associated in the research literature with the risk of suicide. It goes beyond ideation to measure other factors pertinent to suicide risk such as lack of social support, drug abuse and impulsivity. The recruits scored well below the community mean with a mean *T* score of 45 and a low standard deviation of 4.76.

While the average score on suicide measures low, there was one individual scoring more than two standard deviations above the community mean. Morey (1996) uses the mean derived from a clinical normative sample for assessing risk of suicide with elevations of one and two standard deviations representing moderate and marked risk respectively.

The individual in this sample with the elevated score had only a *T* score of 54 relative to the clinical norms and so does not come into the higher risk category as defined by Morey. However, the score is markedly elevated in relation to the community mean. While the score indicates that this person is unlikely to be at imminent risk of self-harm, further investigation would be warranted if this score was obtained from a police applicant or serving officer outside the research context. In the context of this study, it did not appear that the person was at significant risk of harming himself or herself, and a breach of confidentiality could not be justified. In addition, the validity of the individual's protocol is in

doubt as the ICN (Inconsistency) score was elevated enough to possibly indicate a random completion of the test.

8.8 Elevations in Clinical and Treatment Scales

This aspect of analysis was intended to assess the PAI's ability to detect psychological characteristics that might mean the respondent failed to meet the New Zealand Police standard of psychological fitness. This was done by looking at the profiles of recruits who had *T* scores of 70 or greater on one or more full scales.

Using community norms provided by Morey (1996), fourteen recruits had *T* scores of 70 or greater on at least one PAI scale. One of these profiles had a validity scale (ICN) elevated sufficiently to suggest the test was completed randomly and so was dropped from this analysis. The test scores of the remaining thirteen elevated profiles represent a considerable deviation from scores in the community sample and an even wider deviation from law enforcement norms. They are two standard deviations above the mean of the community sample and are in the top 2.2 % of a normal distribution.

The scores are not considered to be necessarily evidence of pathology, but would raise some concerns about those individuals. It is inappropriate to draw conclusions about an individual on the basis of one set of test scores and these should be integrated into as broad a range of other information as possible. It was not possible to gather information such as reports of others to corroborate or disconfirm suggestions from the PAI scores.

Chapter 9: Discussion

9.0 Overview

The selection of suitable personnel for training as police officers or for advanced duties is a crucial function of a law enforcement agency. The particular demands of police work create the need not only to select those who are most suited to the occupation, but screen out those whose psychological characteristics make them unsuitable (Hogg & Wilson, 1995).

The efficacy of using psychological tests for personnel selection has been controversial, but recent research has shown significant relationships between personality and performance (Barrick & Mount, 1991). A number of characteristics have been identified as being valuable for police officers in being able to perform their duties. An absence of significant psychological difficulties has been seen as a fundamental requirement of a prospective police officer (Fyfe et al., 1997), and psychological testing may provide one means of evaluating this.

The New Zealand Police presently utilise a test of cognitive abilities (PL/PQ) and a test measuring aspects of normal personality based on the five factor model of personality (NEO-PI-R). This test battery may lack a test that focuses on the detection of pathological characteristics detrimental to a police career, and the Personality Assessment Inventory has been identified as possibly filling this role.

The present study was a partial replication of one conducted by Black (1995), which assessed the ability of the NEO-PI-R to predict performance of a group of recruits at the Royal New Zealand Police College. The current research intended to assess psychometric properties of the PAI including aspects of reliability and validity, and evaluate the utility of the test as a selection instrument for the New Zealand Police.

The study had a number of aims: (1) to assess the psychometric properties of the PAI using a New Zealand Police sample; (2) to assess the ability of the PAI to predict performance at the Royal New Zealand Police College, and compare this to the NEO-PI-R; (3) to check for possible biases against demographic groups; (4) to evaluate the impact of response sets on test validity; and (5) to evaluate the information provided by the PAI in respect of its ability to screen out those personnel who do not meet the psychological fitness standards of the New Zealand Police.

9.1 Psychometric Properties of the PAI

9.1.0 Reliability

The internal consistency of the PAI was assessed using coefficient alpha and inter-item consistency analysis. Alpha coefficients for PAI full scales were notably lower than for the community sample, with scales measuring Drug Problems, Suicidal Ideation, and Stress performed particularly poorly on this measure. However, most scales did exceed .70, which is often cited as an adequate level of reliability (Schmitt, 1996). No scale reached .85 which Gatewood and Feild (1998) gives as the ideal for a personnel selection measure. The reliability of a test designed to measure psychopathology seems likely to have been attenuated when applied to a population that has already been selected for psychological health.

The Drug and Suicidal Ideation scales in particular seem likely to be unreliable measures in this population. Because of the illegality of drug use and a possibly low base rate of use, police would be unlikely to respond to questions about their use of drugs. Instructions given on the extent of confidentiality may also limit responding on the Suicidal Ideation scale, as respondents may have legitimate fears that endorsing some items may have led to them being identified to police administration.

The PAI subscales demonstrated similar or lower levels of reliability on the coefficient alpha measure compared to the community sample. As a whole, alpha values for the subscales in both samples are reduced as a function of the shorter scale length (Kline, 1993). However, some subscales performed particularly poorly in the police recruit sample such as Health Concerns (SOM-H), Physiological Anxiety (ANX-P), Phobias (ARD-P), and Psychotic Experiences (SCZ-P) with alphas below .5.

The low inter-item correlations found in both the community sample and the police recruit sample demonstrates the heterogeneity of the content being sampled in each scale. However, for some scales in the police recruit sample, inter-item correlation was greatly reduced. This may be a result of the low level of endorsement of items on some scales.

Standard errors of measurement were fairly low, with most below four *T* points for full scales and below five for subscales. However, this possible spread of true scores means that interpretations of observed scores need to be treated with caution. A difference of five *T* points could place a score that appears clinically significant (70 *T* or greater according to Morey, 1991) in the normal range. Boone (1998) suggested that in his sample of psychiatric patients, while most full scales can be interpreted with some confidence, subscale scores should be used for hypothesis generation rather than for making definitive statements. The present study suggests that definitive statements should be avoided for all scales and any hypotheses generated by test scores are used as a basis for further investigation only.

9.1.1 Protocol Validity

The present study found the effect of response sets to possibly be a considerable threat to validity, with only 27.6% of participants being able to be classed as having completed valid protocols with some confidence. However, there were also only about 20% of participants who could be placed in the invalid category with confidence. This is due to the limitation of the validity scales and indexes in terms of their sensitivity and specificity. There remained

almost half the sample whose protocols were classed as needing to be interpreted with caution as a result of possible defensive or random response sets. This indicated that this group of recruits was likely to contain a mixture of valid and invalid protocols.

Based on the work of Peebles and Moore (1998), and Cashel et al., (1995), the "interpret with caution" group should be treated as more likely to be invalid. Both studies found that a lower cutpoint resulted in a more accurate classification of defensive protocols. However, the work of the aforementioned authors is based on studies using college students and inmates, and it is possible that the cutoff points found may not apply to a police sample. All the studies that have been used to develop the validity scales in the PAI have compared scores of a group instructed to fake good with a group instructed to respond honestly. It is not known whether other groups in a real life situation will produce similar scores.

The work of Roberts et al. (1999) has identified elevated PIM scores as common in police samples, and has asserted that it does not usually produce distorted scores on other scales. The lack of collateral information such as clinical interviews or reports of others means that it is not possible to confirm that a recruit is experiencing psychological difficulties which he/she has chosen not to report in the test situation. This leaves the question of whether the evidence of defensive responding found in the recruit sample is to be believed.

Protocols classed as invalid produced substantially lower scores on PAI clinical and treatment scales. There are several hypotheses for these results. Taken at face value, it appears that the defensive responding indicated by higher scores on scales such as PIM (Positive Impression Management) has suppressed responding on other scales. It is possible that the respondents in the invalid group have been reluctant to endorse items that are socially undesirable such as those measuring schizophrenic symptoms and that their true scores may be higher.

An alternative interpretation of the data is that respondents in the invalid group are more psychologically healthy than the valid group. The PIM scale attempts

to measure defensiveness by asking respondents to admit to minor flaws common to the general population. Respondents may be answering honestly (but possibly unrealistically) to these items, and are not deliberately attempting to create an overly favourable impression. The factor elevating PIM may be an unrealistic appraisal of self, as described by Paulhus and Reid (1991), rather than deliberate attempts at impression management.

The PAI validity scales provide a means of assisting the interpretation of not only the PAI scores, but also possibly other self-report data such as the NEO-PI-R. The findings of the present study brings into some doubt the wisdom of assessing personality based on a single test such as the NEO-PI-R which does not assess for response sets. The authors of this test make a case for not including validity scales that includes the possible alienation of respondents who may perceive the validity related items as irrelevant or silly (Costa & McCrae, 1983), and the inability of validity scales to reliably distinguish between valid and invalid protocols (Costa and McCrae, 1992b).

However, the interpretation of self report personality inventories rests on integration of test scores with other sources of information about the respondent such as background checks and reports of others (Morey, 1996). It seems sensible to also integrate validity scale information, as another source of data with which to interpret test scores.

It is not possible to say with a high degree of certainty if a respondent has completed a test in an honest careful manner. It would appear to be logical for a police applicant to not endorse items on scales such as drug use, suicidality or psychosis if they were motivated to gain entry to the police. As Rosse et al. (1998) point out, it would be surprising if most job applicants did not positively distort some answers, given the demand characteristics present in the testing situation.

A sophisticated respondent may be able to present in an overly positive or negative fashion, without being detected, if they do it in a believable manner (Cashel et al., 1995). Many PAI items are obvious in the content they are

attempting to measure, and respondents may be able to manage their responses on critical items to avoid disclosing pathology without elevating validity scales.

Despite this, the more subtle nature of items on the Positive Impression Management scale in particular may catch out many respondents who are not aware of the nature of such scales. Efforts to present in an unrealistically positive manner by failing to admit to minor faults then suggest that responding to items on more obviously pathological items is likely to be suppressed (Morey, 1996). This does not rule out the possibility, however, that the respondent has no significant psychopathology, and simply has an unrealistically positive appraisal of his/herself.

The use of validity scales has significant limitations in terms of what inferences can be drawn from the scores obtained, but does provide another source of information with which to interpret test scores. The poor sensitivity and specificity qualities of the validity scales make judgements of protocol validity difficult. The assertion by Roberts et al. (1999) that elevated PIM scores does not suppress responding in police samples casts further doubt on the utility of attempting to assess defensive responding in this population.

9.1.2 Effects of Response Sets on Predictive Validity

Most studies have found that response distortion does not significantly affect the ability of personality tests to predict job-related criteria (Barrick & Mount, 1996; Ones, Viswesvaran, & Reiss, 1996; Hough, 1998; Hough et al., 1990). The current study supports these findings, as the regression analyses showed that protocol validity did not significantly affect the ability of either test to predict performance at the RNZPC. The correlations produced were low and none reached a significant level.

A mixed pattern of results was revealed in the comparison of univariate correlations between scale scores and performance. A comparison was made between correlations between PAI test scores and final grade obtained at the

RNZPC for valid, invalid and all protocols. It was found that for some scales, invalid protocols produced stronger correlations with performance than for all protocols together. The valid PAI protocols produced fewer significant correlations, which may have been influenced by the low number of protocols in this group, which will have reduced the statistical power of the analysis. However, those correlations found were higher than for the other categories of protocols.

Comparisons for the NEO-PI-R showed the invalid protocols producing only one significant correlation, but this was slightly higher than when all protocols were included in the analysis. The valid protocols produced one significant correlation that was considerably higher than any other.

Protocol validity does appear to have an impact on univariate correlation between scores on the personality tests and performance. The low *N* in the valid group creates difficulties in drawing inferences from this analysis, but it does appear that relationship between test scores and performance is stronger in the group classed as unaffected by response distortion.

Despite some possible influence on univariate correlation between personality and performance, the issue of response distortion does not appear to be a significant factor influencing predictive validity. However, the impact of response distortion may be more pertinent when making clinical interpretations of test scores. The PAI is intended to function as a broad measure of psychopathology, and rests on the interpretation of clinically significant scores. Generally *T* scores of 70 or greater (using community norms) are considered significant in terms of revealing probable pathology in the respondent (Morey, 1991). Suppression of clinical scale scores by even a relatively small margin caused by response distortion could significantly influence the accuracy of interpretation of test scores.

A comparison of *T* scores between invalid and valid protocols for the PAI showed significant differences for most scales with valid protocols scoring as

much as six *T* points higher. This amounts to over one half of a standard deviation, and could result in a scale score failing to reach clinical significance.

In summary, the influence of response sets on the use of PAI test scores is primarily an issue when making judgements about psychological issues indicated by elevated scales. Protocol validity does not seem to be a significant problem when using scores to predict performance.

9.1.3 Construct Validity

The construct validity of the PAI was investigated by correlating PAI scale scores with each other and the NEO-PI-R. The relationships between theoretically similar and dissimilar scales were in the expected direction showing that the PAI does have some degree of convergent and discriminant validity in this population. However, the ANX (Anxiety) and ARD (Anxiety-Related Disorders) scales showed some lack of discriminant validity in correlating equally with the NEO-PI-R measures of depression and anxiety.

The strength of the relationships between PAI and NEO-PI-R scales was generally lower than that found by Morey (1991) and Cashel et al. (1992b) in other populations. The reason for this was not established, but it seems likely that the distorting effects of defensive responding, and the restricted range of scores found in this highly selected population may have attenuated correlations.

9.1.4 Structure of the PAI

Principal component analysis of the PAI correlation matrix revealed four factors. Factor 1 (representing neuroticism and psychological distress) and Factor 3 (representing social detachment and interpersonal difficulty) are similar to those found by Morey (1991). Factor 2 (representing mania, aggression, dominance and disregard for others) is similar to Morey's Factor 3 except for the loading of ANT (Antisocial Features) on the factor in the present study. Factor 4 was marginal for inclusion in the analysis and appears to represent some aspect of

infrequent responding as it contains two scales that were endorsed at very low levels in the recruit sample.

While there was some similarity between factors extracted from the police sample and Morey's community sample, elements of mania, aggression, dominance and antisocial features (Factor 2) accounted for more variance in the present study. This aspect of analysis was intended to provide a further description of the scores in the present sample. It appears that despite using the same factor analytic technique as Morey (1991), some significant differences in the factor structure of the PAI were found. This is hardly surprising given the differences between Morey's sample (a census-stratified sample of the North American population), and the sample in the present study (a highly selected group of New Zealand police recruits).

9.1.5 Criterion-Related Predictive Validity

PAI and NEO-PI-R scales produced similar levels of correlation with final grade, although fewer NEO-PI-R scales reached significance. Black's (1995) study with another sample of New Zealand police recruits found significant correlations on many facet scales that did not reach significance in this study. Black found correlations on facet scales of all five personality domains, with all Conscientiousness facet scales having significant correlations. However, Black's sample was considerably larger (N = 238) compared to the sample in this study (N = 106), and it is possible that correlations were attenuated by the smaller sample size.

The strength of correlations between test scores and performance was similar in the present study with that found in police samples by Black (1995) and Barrick and Mount (1991). The findings of the present study appear to differ from the aforementioned ones, in that negative psychological characteristics such as those measured on the NEO-PI-R Neuroticism scales, and the clinical scales of the PAI have correlated as highly with performance as positive characteristics such as conscientiousness. For example, the correlation

between the Neuroticism facet scale N5 (Impulsiveness) and performance was higher in the present study (r = -.26) compared to Black's study (r = -.17).

The finding that N5 (Impulsiveness) correlates negatively with performance concurs with Costa and McCrae (1995). They found that police applicants who were recommended as suitable based on an interview by psychologists scored lower on this scale. However, the other predictors found in the present study (O5: Ideas & C4: Achievement Striving) did not agree with Costa and McCrae's findings (N2: Angry Hostility; N6: Vulnerability; C1: Competence; C5: Self-Discipline).

The PAI scales DEP (Depression) and ANX (Anxiety) had a significant negative relationship with performance. The characteristics measured by these scales may have interfered with performance. Those experiencing even mild depressive symptoms may have been less likely to study and anxiety may have interfered with performance during assessment.

The only PAI scale to have a significant positive association with performance was DOM (Dominance). Morey (1991) describes those scoring higher on this scale as being self-assured, confident, and forceful. These would seem to be qualities that would assist a person in an environment such as a police training program.

There may be a higher order characteristic explaining some of the variance in performance. Two of the PAI scales that correlate significantly with performance are those such as SOM (Somatic Complaints) and SUI (Suicidal Ideation), which had very low levels of endorsement. Those who did endorse items on these scales seem likely to differ considerably from the rest of the sample in some way.

9.1.6 Prediction of RNZPC Final Grade with the PAI and NEO-PI-R

The regression analysis performed found that the PAI full and subscales did predict performance, but only accounted for 12% of variance in the total grade

achieved at the RNZPC. The theoretical relationship between individual scales and performance was unclear with only the DOM (Dominance) full scale and PAR-P (Paranoia) subscale achieving significant correlations in the analysis.

It is notable that although scales were selected for inclusion in the analysis on the basis of their univariate correlation with performance, many did not appear to have a significant contribution in multivariate analysis. This is likely to be due to a substantial overlap between scales in the variance in performance that they explain (Tabachnik & Fidell, 1989).

The NEO-PI-R facet scales included in the regression analysis together accounted for 8% of the variance in final grade. No scale reached significance on it's own. This was also the case for the NEO-PI-R domain scales, which accounted for only 5% of variance in final grade.

The incremental validity of the PAI in the New Zealand Police test battery was assessed using regression analysis, with NEO-PI-R scores in the first step and PAI scores in the second. This looked at what increment in predictive ability was gained by including the PAI in the equation. The addition of PAI full scales resulted in a significant change in R^2 of .13. In this analysis, the NEO-PI-R did not contribute significantly. This may have been due to the removal of cases that were outliers on the PAI scales, which changed the sample from that used when the NEO-PI-R was assessed on its own. Similar results were obtained with the PAI subscales, although the R change was only borderline significant.

The results of regression analysis showed that PAI scores accounted for more variance in final grade than the NEO-PI-R, although neither test produced more than a very modest effect. It seems debatable how useful a test is for making selection decisions on the basis of accounting for only 12% of variance. However, when the NEO-PI-R was included in the New Zealand Police test battery along with a cognitive abilities test, it produced a substantial increment in predictive validity (Black, 1995). It seems likely that the PAI would produce a similar result and so may be a useful addition to the battery. However, the findings of regression analyses show that little or no increment in validity was

gained by using both the PAI and NEO-PI-R together compared to the PAI on its own.

The findings of this study conflict with the body of evidence suggesting conscientiousness to be the characteristic most predictive of performance (Barrick & Mount, 1991). It appears that for some reason, negative characteristics measured by both the PAI and NEO-PI-R was at least as strongly correlated with performance as conscientiousness. The relatively low sample (N = 127) size may have hampered efforts to find significant correlations.

9.2 Test Fairness

9.2.0 Age

The relationship between age and test scores was evaluated using Pearson correlations. Age correlated negatively with a range of clinical PAI scales and subscales indicating that increasing age was correlated with lower PAI scores. All correlations were low with none over .30. However, Morey (1996) identified age as an influence on several scales and the ANT (Antisocial Features) scales in particular. In the present study, the ANT scales featured in the profiles of those who had clinically significant scores (T > 70). With most recruits being under 30 years of age, an association between age and test scores needs to be considered when interpreting such scales. For practical purposes, this consideration might simply consist of placing less emphasis on isolated elevations on age related scales in younger recruits.

Only one NEO-PI-R scale correlated significantly with age. This was for the E5 (excitement-seeking) facet scale, which indicated that increased age was correlated with a decrease in excitement-seeking. This concurred with the PAI's ANT-S (Stimulus Seeking) scale also correlating negatively with age. It is not surprising that an increased desire for excitement was measured in younger recruits by both tests.

9.2.1 Gender

On PAI scales, a number of significant differences were found between genders with males scoring lower than females in almost all cases. The exception was the AGG-P (Physical Aggression) subscales where males scored higher. Most of the significant differences were found on scales measuring some form of emotional disturbance such as anxiety. Kaplan and Sadock (1998) state women are more likely to develop an anxiety disorder during their lifetime (30.5%) compared to men (19.2%). Other differences between genders on the PAI are consistent with reported epidemiological data (Kaplan & Sadock, 1998). The slightly elevated score on MAN-A (Activity Level) concurs with the finding of Carpenter and Raza (1987) that female police applicants had higher energy levels. On most scales, the difference between genders is small with almost all being three *T* points or less.

The number of scales with significant differences between genders was greater than that reported by Morey (1996) who found only one (Antisocial Features). The low number of women (N = 38) in the sample in the present sample may have influenced this finding, and it is possible that the personality characteristics of women attracted to a police career may differ considerably from the wider population. Another possibility is that the males in the sample differ significantly from men in the community. The differences found between genders do not appear large enough to suggest that the test may be biased against men or women, or that separate norms based on gender are required.

Some larger significant differences were found between genders on the NEO-PI-R with three of the four differences exceeding five *T* points. Males scored higher on A1 (Trust) and C2 (Order), while females scored higher on O1 (Fantasy) and O6 (Values). Costa and McCrae (1992a) report higher scores for women on N and A facet scales, but that differences are usually small. The New Zealand Police use norms for the NEO-PI-R that have been developed in the police population and have separate norms for males and females.

9.2.2 Ethnicity

Three significant differences were found on PAI full scales with Maori scoring higher than Caucasians on SOM (Somatic Complaints), DEP (Depression) and WRM (Warmth) with differences of between four and five *T* points. On PAI subscales, Maori scored higher on SOM-S (Somatisation), SOM-H (Health Concerns) and DEP-P (Physiological Depression), with differences ranging between two and five *T* points. Caucasians scored higher on AGG-V (Verbal Aggression) by five *T* points.

The low number of Maori in the sample makes it difficult to draw conclusions from the results found. However, the elevated subscales for Maori related to physiological experiences suggest there could be a difference in the way somatic phenomena are experienced or perceived by Maori. The elevations on Somatic Complaints and Depression may be related, as Maori are more likely to present with vague somatic complaints when depressed (National Health Committee, 1996).

The difference on AGG-V suggests the group of Maori in the sample were less likely than Caucasians to express aggression verbally. Without corroborating or disconfirming evidence, it is not known whether the differing scores found between ethnic groups are indicating actual differences in the constructs being measured, or a difference in the way the test items are being perceived and responded to.

On NEO-PI-R facet scales, Maori scored higher on N3 (Depression), N5 (Impulsiveness), O2 (Aesthetics) and A6 (Tendermindedness) with differences ranging between four and seven *T* points. The same points made in interpreting the differences in PAI scoring are applicable here also. If the differences in test scores do represent actual differences in the constructs measured, it appears that the Maori recruits in this sample are more likely to experience depressive affect, have a higher appreciation of art and beauty, and have more of a sympathetic and concerned attitude toward others.

There seems to be some concurrence between the PAI and NEO-PI-R in the differences found between ethnic groups. In particular, both tests have found Maori to report more depressive symptoms (DEP and N3), and to perhaps have a different interpersonal style (WRM and A6) that is more empathic and concerned for others. The difference found in interpersonal style may reflect a more collective cultural orientation, where interdependence rather than independence is valued (Durie, 1994).

The conclusions drawn from this analysis are limited to police recruits as persons of any ethnicity who have joined the police may differ considerably from the general population. Without other information, it is difficult to determine if the differences found between demographic groups using PAI and NEO-PI-R scores indicate actual differences in personality, or are due to other factors such as differing perceptions of test items and consequent variations in responding.

9.2.3 The Four-Fifths Rule

The four-fifths rule dictates if the selection rate for any demographic group is less than four-fifths of the group with the highest selection rate, this may be considered evidence of unfairness. This aspect of analysis used the assumption that in a selection situation, those with *T* scores exceeding 70 on PAI scales would be screened out. This assumption is unlikely to be true in a real selection situation, as a great deal of other information is utilised when making decisions about an applicant. It was made, however, in order to allow an approximate application of the four-fifths rule to see if there may be a possibility of the PAI having an adverse impact on a demographic group.

While there was little difference in selection ratios between genders, the selection rate for Maori was only 76% of that of Caucasians. This resulted from a disproportionate number of Maori in the group who had elevated PAI scale scores. The scales found to have significant differences in scoring between Maori and Caucasian using *t* tests did not correspond to the scale (ANT; Antisocial Features) most commonly exceeding 70 *T* for Maori. It appears that while as a group, there was little difference in PAI scores compared to those of

Caucasians, there was a subgroup with high ANT scores who may differ in some way from the other Maori in the sample, and from the Caucasian group.

The results of this analysis do not constitute evidence of adverse impact on Maori, but do raise the possibility of this. It is not possible to determine if Maori are over represented in the elevated subgroup because of real individual differences, or resulting from a group difference in how Maori respond to certain test items. Because Maori were not significantly different from Caucasians as a group on the ANT scale, it seems likely, however, that the test may have picked up true individual differences in those Maori who were highly elevated on this measure.

It would be informative to check on the contribution of historical issues to scales such as ANT-A (Antisocial Behaviours), as Maori are more likely to have had some contact with the legal system (McFarlane-Nathan, 1999). Groth-Marnat (1997) suggests it is possible that differences found between ethnic groups may in fact be true differences caused by the increased stresses experienced by minorities. In this case, it would seem appropriate to follow the advice of Graham (1987, cited in Groth-Marnat, 1997). In regard to the MMPI, he suggests that clinicians should tentatively accept deviant scores, but take particular care to integrate other information with the test scores.

Boyle and Lennon (1994) raised the issue of the potential offensiveness of certain PAI items. Representatives of North American cultural groups reviewed the items in order to screen for culturally biased or offensive items (Morey, 1991). It would be useful to carry out a similar process with New Zealand groups and in particular Maori. Although items cannot be eliminated, an understanding of how particular ethnic groups view the content of the PAI would be informative in deciding on the appropriateness of the test for the New Zealand situation and also when interpreting scores.

9.3 PAI Scores for the Police Recruit Sample

The comparison of PAI scores for the police recruit sample with the normative community sample revealed that on most clinical and treatment scales, the recruits produced lower *T* scores. If the validity of the protocols were assumed, this would indicate that the recruit sample had a higher level of psychological health than the community sample. The scales ANT (Antisocial Features), ALC (Alcohol Problems), and RXR (Treatment Rejection) were elevated. ALC was only marginally higher however. On PAI subscales, most scales were below the community mean. ANT-S (Stimulus Seeking) stood out, as being considerably higher as it was nearly one standard deviation above the community mean.

Unfortunately, the raw scores of the sample in this study were not able to be normed against police samples as well as the community sample. The normative information for the police sample is considered proprietary information by the test publisher (Psychological Assessment Resources), and can only be accessed through the purchase of an interpretive software package (PAI-SP: PAI Law Enforcement, Corrections, and Public Safety Selection Report Module).

Some comparisons could be made, however, as the manual (Roberts et al., 1999) for this software contained the profiles of several police applicants with their scores normed against both community and police groups. What was apparent from these profiles was that on most clinical and treatment scales, raw scores that are around the mean for the community sample may be one or two standard deviations above the mean of the police normative sample.

The most highly elevated clinical scale, ANT (Antisocial Features), is likely to have been influenced by the age of participants, as 70% of the recruits in the present study were under 30 years of age. Morey (1996) found that respondents in the 18 to 29 year old age bracket scored up to seven *T* points higher than the community as a whole.

However, according to Roberts et al (1999), the ANT full scale was not normally elevated in a North American law enforcement sample, and a *T* score of 56 (normed against the community sample) becomes 67 *T* when compared to the public safety norms. This suggests that the high ANT score in the sample in the present study may be unusual for a police sample.

The recruit's elevation on the subscale ANT-S (Stimulus Seeking) is likely to also be at least partly a result of the age of participants, as the mean of 18 to 29 year olds is 57 *T* (Morey, 1996). An elevation on the ANT-S scale is not considered to be pathological on its own, but is likely to indicate problems if it is paired with other traits such as poor impulse control or anger problems. This is because those scoring high on ANT-S may lack the inhibiting effects of anxiety, which normally serves to place some limits on behaviour that is potentially harmful to self and others.

A feature of the police recruit scores was the reduced variance in scoring, reflected in low standard deviations that were often less than two-thirds of that of the community sample. This was expected, given that the recruits have been screened extensively prior to entering training, and those with gross psychopathology should have been removed. Despite the low variance in scores as a group, almost every scale had high scores that were more than two standard deviations from the community sample, and even more extreme compared to the police recruit mean. It is these elevations that are discussed in a later section as they represent such a considerable deviation from the responding of the recruits as a group.

9.4 Potential Danger to Self and Others

The purpose of this analysis was to focus on aspects of the PAI specifically designed to assess a respondent's potential to harm him/herself and others. This investigation took a narrow focus on harm-related scales, but in the context of a normal assessment, the respondent's entire PAI profile would be evaluated.

The recruit sample was assessed for potential for aggression using the AGG scales and the Violence Potential Index derived from relevant aspects of the PAI profile. While no recruit was classified by Morey's (1996) guidelines as being a significant risk for violent behaviour, one person had a score found in the range of a sample of persons with assault convictions.

The AGG subscales allowed the different ways in which aggression is expressed to be measured. As a group, the recruits scored below community norms on AGG-A (Aggressive attitude) and AGG-P (Physical aggression), but higher on AGG-V (Verbal aggression). It is possible this is linked to a desirable component of a police officer's personality, which is the ability to express him/herself in a forceful manner when required (Black, 1995).

Potential for self-harm was assessed using the SUI (Suicidal Ideation) scale and the Suicide Potential Index, which is derived from relevant aspects of the PAI profile. The group of recruits produced a very low mean score on this measure with only one person raising concern. This person's score was highly elevated against community norms, but close to average against clinical norms. On the basis of the information provided by Morey (1996), the score did not warrant confidentiality being breached. Under selection circumstances however, it seems likely that this person would be investigated further.

The specific information provided by the PAI relating to potential harm to self and others seems useful for the purposes of selection, and also for ongoing monitoring of serving officers health. There remains the difficulty of relating this information to the low frequency of self-harming behaviours in particular.

Between 1990 and 1997, 12 officers committed suicide representing 37.5 % of all deaths in this period (New Zealand Police, 1998). The police suicide rate is only slightly above the national rate (20.1/100,000/pa compared to 17.1/100,000/pa) and the small sample size makes this comparison difficult. However, the rate of suicide in police is of concern to the organisation (New Zealand Police, 1998), and any information that would contribute to prevention would seem to be useful.

In the context of police work where the use of force is commonplace in controlling offenders, information regarding aggressiveness may be of some use in identifying officers who may be prone to inappropriate use of force. Further investigation is required to determine if elevated scores on the PAI aggression measures is related to the incidence of misuse of force.

9.5 Elevations in PAI Clinical and Treatment Scales

Using community norms and guidelines provided by Morey (1996), thirteen recruits scored at 70 or greater on PAI scales. Scores such as this represent a considerable deviation from scores in the community sample and an even wider deviation from law enforcement norms. The scores are not considered to be necessarily evidence of pathology, but would raise some concerns about those individuals. It is inappropriate to draw conclusions about an individual on the basis of one set of test scores, and these should be integrated into as broad a range of other information as possible.

However, the PAI is constructed with obvious items that are designed to directly measure symptomatology of recognised clinical disorders, and endorsing items on these scales does suggest that the person may be experiencing some difficulties. It is interesting that the PAI has identified a relatively large number of persons in a highly selected sample as possibly having some psychological issues.

There are some possible explanations for this finding, although in the absence of other information, no answers can be given with any confidence. The first possibility is that each of the thirteen recruits does have the psychological problems suggested by the test scores. Another is that test items are responded to in a different fashion by a sample of New Zealand police recruits from that of the North American community sample, and that the recruits do not have the problems suggested by the PAI scores.

The test items were selected on the basis that their content reflects the phenomenology of the mental disorder being measured. This was done using a panel of experts who are recognised in the assessment of constructs measured by the PAI. The rational method of item selection, as opposed to external methods of test construction, should help to overcome the problem of test items being specific to the population against which they were developed (Morey, 1991). Directly measuring symptomatology that is accepted as being indicative of a particular disorder enables PAI test scores to be more easily interpreted by clinicians. Interpreting the PAI does not require reference to obscure past studies, or a complicated transformation of scores, as is often the case for the MMPI (White, 1996).

Despite the unexpected number of elevated profiles, few suggest that the recruits are suffering from a major psychological illness. The most common full scale elevation was on the ANT (Antisocial Features) scale, which measures aspects of personality and behaviour pertinent to the constructs of antisocial personality and psychopathy. This was thought to have been possibly influenced by the age of participants as the mean *T* score for persons under 30 was seven *T* points higher than the community mean (Morey, 1991).

Examination of common PAI subscale elevations in this sample revealed a scale (ANT-S; Stimulus Seeking) that may be elevated as a function of the personality of people attracted to police work and may not indicate pathology. Eber (1991, cited in Lorr & Strack, 1994) found that police applicants were high on measures of thrill seeking and disregard for social conventions. Aylward (1985) describes speculation around whether some personality characteristics

in police personnel may be so different, that what is considered atypical in the general population is normal in a police officer.

Seven of the elevated profiles had ANT-S T scores of 70 or more. The items contained in this subscale involve the enjoyment of danger, risk taking, and fast driving, which would seem to fit the image of the new police officer who is attracted to the excitement of job aspects such as chasing and apprehending offenders. Other aspects are less orientated to job characteristics such as reporting wild behaviour and not being inclined to settle down in one place. It is not known if there was differential responding on perhaps job orientated aspects of this scale.

Elevations of a scale such as ANT-A (Antisocial Behaviour) seem to be a contraindication for a police career, as the items tap past and present unlawful and antisocial behaviour. Five recruits were highly elevated on this subscale, and it is difficult to find an alternative explanation for these elevations, as the items are unambiguous and ask about events and behaviours. Evidence of even historical antisocial behaviour is a serious issue when selecting police, as past behaviour is a predictor of future behaviour (Sarchione, Cuttler, Muchinsky, & Nelson-Gray, 1998). It is possible, however, that the historical issues that may have contributed to elevations were investigated by police during the application process, and were considered to be insignificant in terms of the applicant's recent behaviour and suitability for police work.

It is interesting to note the findings of Roberts et al (1999) who found that of those who scored 60 *T* or higher (using police norms) on the ANT-A scale, 26% admitted to illegal drug problems, 36% to alcohol use problems, 62% to anger management problems, and 55% to job related problems. Similar results were found for those who elevated on the full ANT scale. Because the scores were normed against the police sample, these respondents would have produced lower *T* scores with community norms.

In addition to the elevations on ANT scales, some recruits elevated on other clinical and treatment scales, including ALC (Alcohol Problems), MAN (Mania),

STR (Stress) and ANX (Anxiety). Respondents in Roberts et al. (1999) sample who elevated on these scales reported similar or lower levels of problems to those scoring highly on the ANT scale.

The question of whether the elevations found reflect actual pathology needs to be answered if the PAI is to be used for the New Zealand Police. The integration of the PAI data with other sources of information should go some way toward verifying its validity. In a selection situation, persons whose scores reached clinical significance would be likely to be referred to a clinical psychologist for a comprehensive evaluation, as has been recommended by the Police Education Advisory Council (1998).

While it is debatable if any of the recruits with elevated PAI scores would actually be diagnosable with a clinical disorder, their scores are markedly higher than for the community and police normative groups. These people may differ in some way from other recruits that may be important for their future performance as police officers. What is of greater concern, however, is that the criteria used to identify elevated profiles are, according to Roberts et al (1999), highly conservative. The application of norms produced from North American police would have resulted in many more recruits being identified as possibly having problems relevant to a police career.

9.6 NEO-PI-R vs PAI

Ross, Bailley, and Millis (1997) found that the NEO-PI-R is highly susceptible to positive self-presentation. The apparently high level of defensive responding found with the PAI suggests that the NEO-PI-R may also be influenced by this response set. The NEO-PI-R does not include any information with which to internally check the validity of the test protocol with respect to defensiveness. The PAI with its validity scales appear to provide some advantage over the NEO-PI-R in this area, but has the limitations in accurate detection mentioned previously.

The NEO-PI-R has been included in the New Zealand Police test battery primarily as a screen in tool to select personnel with desirable psychological characteristics rather than to screen out unsuitable applicants. Using the criterion of performance during training, it was shown to be less effective in its predictive ability than the PAI.

The effectiveness of the screening procedures has been questioned by the number of recruits scoring at very high levels on the PAI. Although no firm conclusions can be drawn on the basis of one test, it appears that the inclusion of a measure focused on assessing pathological characteristics might enhance the effectiveness of the test battery.

The PAI has provided information that, under selection circumstances, would be useful in identifying applicants who required further investigation. The fact that the test did so in an already highly selected population suggests that administering the test to an unscreened applicant group would result in even higher rates of possible problems identified. However, the NEO-PI-R profiles for this sample were not interpreted which leaves the possibility that the test did identify applicants with possible psychological problems, but that the person was selected on the basis of other disconfirmatory information.

9.7 Limitations

A number of limitations with the current study should be noted. What appears to be the most significant limitation of the study is the lack of other information with which to corroborate any interpretation of test scores. In the absence of other evidence of pathology, an elevated scale score is simply a number.

Another limitation is the lack of job-related criteria to relate to scores obtained on the PAI and NEO-PI-R. The use of data from the RNZPC did provide useful information on the ability of the tests to predict performance on academic and practical measures. However, the abilities measured at the RNZPC may not generalise to the work situation.

The recruits in each wing were asked to participate and a significant number chose not to take part in this study. It is not known if the non-participants differed in some way from those who did participate. It is possible that a refusal to be part of a study of this nature may indicate an increased level of defensiveness in being reluctant to undergo a psychological evaluation, however limited that may be. The nature of questioning from a few participants would indicate that this may have been the case for some, but many may have simply not wished to spend an hour completing a test of any kind. This is particularly relevant for the groups who were tested during their free time. It was preferable, however, that those who were reluctant to take part were encouraged not to participate, as such a person may not have answered the test questions in a straightforward manner.

The level of defensive responding found by PAI scales such as PIM (Positive Impression Management) indicated that this might have compromised the validity of much of the data. It is not possible to predict which scales may have been affected by the response set. Morey (1996) believes that defensiveness is unlikely to have a global effect in suppressing scores across all scales, and will particularly impact on specific scales. It is not known which scales these may be.

It may have been preferable to use different instructions when administering the test. Rather than simply requesting honest responding, it may have been useful to mention the ability of the test to detect defensive or random responding. It is debatable however whether this would reduce defensiveness or simply encourage respondents to fake good in a more believable fashion. Baer and Wetter (1997) found that providing information about the presence of validity scales resulted in increased ability to fake good without elevating the PIM scale.

The study was limited by the use of norms based on a community sample rather than police. Roberts et al. (1999) use of the PAI with police suggests that normative scores with this group differ considerably from the community. In

particular, the PIM scores used to evaluate defensive responding in this study may be incorrect for police.

The study used police recruits who had already gained entry to the police and for whom the test scores would not affect their position in the police. It is unclear how well results from this group would generalise to an applicant sample, either for those applying to join the police, or for incumbent police applying for special duties. Cortina et al. (1992) also identified the use of recruits rather than applicants as a significant limitation of their study.

9.8 Directions for Future Research

Defensive responding has been identified by this study as a significant threat to validity in this population and investigating methods of reducing the effect of response sets would be a useful direction for future study. This might involve looking at the effect of differing test instructions relating to honest responding or developing a discriminant function to detect defensiveness in this population as was done for students and inmates by Cashel et al. (1995).

The lack of specifically job related criteria hampers the investigation of criterion-related validity. A more useful measure of this aspect of validity may be gained by following up the participants in the present study at a later date and assessing them on a more relevant basis than academic performance. Some possible measures might include disciplinary actions, length of service, and referrals to counselling.

Another avenue for investigation would be to readminister the PAI in the future. This would be of interest in seeing if any significant changes occurred in PAI scales scores over time, and relating these to hypothesised effects of police work such as increases in stress and anxiety. It is also important to evaluate test-retest reliability for this population.

It would be useful to examine responding on the PAI at an item level, and compare responses to normative groups. There may be some scales which are elevated by idiosyncratic responding with this sample such as has been suggested for the ANT-S subscale.

The utility of community based norms needs to be investigated as Roberts et al. (1999) raises significant questions about this issue. It may be necessary to use the law enforcement norms collected by Roberts, or develop a New Zealand normative group in order to interpret the PAI in a more meaningful way for police selection.

It is not known at this point if elevated scores on the PAI will be expressed as disordered behaviour in the same manner as that found by Morey (1996) with the North American population. Further investigation is required to match PAI scores with other information in order to make some judgements about this issue.

9.9 Conclusions

The psychometric properties of the PAI appear in most cases to be adequate, although some scales demonstrated lower levels of internal consistency than is desirable. Assessment of construct validity showed that for most scales, the PAI correlated in a theoretically concordant manner within itself and with the NEO-PI-R.

The issue of response distortion has not been fully clarified by this study. Based on community norms and guidelines produced by the test author, the sample contained a high number of respondents who were attempting to create an unrealistically positive impression. This did not affect the ability of the test to predict performance however. It is not clear how best to interpret protocol validity information in a screening context, as there is a suggestion that elevated defensiveness scores are normal in this population and does not reflect distortion.

An investigation into the possibility of the PAI having adverse impact on demographic groups showed that while there was little difference in scores based on gender, Maori and Caucasians differed significantly on some scales. Of more concern was the disproportionate number of Maori making up a subgroup of all participants who had highly elevated scores. It was not able to be determined if this resulted from some form of test bias, or the accurate measurement of individual differences.

The PAI has demonstrated modest predictive ability using the final grade achieved during training as a criterion. The validity coefficients achieved however were significantly higher than for the NEO-PI-R, and comparable to other studies using personality characteristics to predict performance.

The PAI shows promise as a tool for screening for psychological characteristics detrimental to police work. It has produced information that could be utilised to prompt further investigation of some participants if they were in a selection context.

References

Adlam, K.R.C. (1982). The police personality: Psychological consequences of being a police officer. *Journal of Police Science and Administration, 10(3),* 344-349.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.).* Washington, DC: Author.

Anastasi, A., & Urbina, S. (1997). *Psychological testing.* (7th ed). New Jersey: Simon & Schuster.

Anson, R.H., & Bloom, M.E. (1988). Police stress in an occupational context. *Journal of Police Science and Administration, 16(4),* 229-235.

Aylward, J. (1985). Psychological testing and police selection. *Journal of Police Science and Administration, 13(3),* 201-210.

Baer, R.A., & Wetter, M.W. (1997). Effects of information about validity scales on underreporting of symptoms on the Personality Assessment Inventory. *Journal of Personality Assessment, 68(2),* 402-413.

Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1-26.

Barrick, M.R. & Mount, M.K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81(3),* 261-272.

Bartol, C.R. (1991). Predictive validation of the MMPI for small-town police officers who fail. *Professional Psychology: Research and Practice, 22(2),* 127-132.

Bartol, C.R. & Bartol, A.M. (1994). *Psychology and law* (2nd ed.). Belmont, CA: Brooks/Cole.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *BDI-II Manual.* San Antonio, TX: The Psychological Corporation.

Bennett, L.A. (1990). The untapped potential of psychological assessments. *The Police Chief, February, 23-25.*

Ben-Porath, Y.S. & Waller, N.G. (1992). "Normal" personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment, 4(1),* 14-19.

Beutler, L.E., Storm, A., Kirkish, P., Scogin, F., & Gaines, J.A. (1985). Parameters in the prediction of police officer performance. *Professional Psychology: Research and Practice, 16(2),* 324-335.

Beutler, L.E., Nussbaum, P.D., & Meredith, K.E. (1988). Changing personality patterns of police officers. *Professional Psychology: Research and Practice, 19(5),* 503-507.

Black, J. (1995). *Psychological testing project: Selection for success. Recruit psychological testing project: Principles of psychological testing and summary of research findings.* Wellington, New Zealand: New Zealand Police.

Blau, T.H. (1994). *Psychological services for law enforcement.* United States: John Wiley & Sons.

Boccaccini, M.T., & Brodsky, S.I. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Practice: Research and Practice, 30,* 253-259.

Boone, D. (1998). Internal consistency reliability of the Personality Assessment Inventory with psychiatric inpatients. *Journal of Clinical Psychology, 54(6),* 839-843.

Boyle, G.J., & Lennon, T.J. (1994). Examination of the reliability and validity of the Personality Assessment Inventory. *Journal of Psychopathology and Behavioral Assessment, 16(3),* 173-401.

Boyle, G.J., & Lennon, T.J. (1996). Psychometric limitations of the Personality Assessment Inventory: A reply to Morey's (1995) rejoinder. *Journal of Psychopathology and Behavioral Assessment, 18,* 197-204.

Burbeck, E.L. & Furnham, A. (1985). Police officer selection: A critical review of the literature. *Journal of Police Science and Administration, 13,* 58-68.

Burkhart, B.R. (1980). Conceptual issues in the development of police selection procedures. *Professional Psychology,* 121-129.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J.N. & Rouse, S.V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47,* 87-111.

Butcher, J.N., Morfitt, R.C., Rouse, S.V., & Holden, R.R. (1997). Reducing MMPI-2 defensiveness: The effect of specialised instructions on retest validity in a job applicant sample. *Journal of Personality Assessment, 68(2),* 385-401.

Carpenter, B.N., & Raza, S.M. (1987). Personality characteristics of police applicants. *Journal of Police Science and Administration, 15(1),* 10-17.

Cashel, M.L., Rogers, R., Sewell, K., & Martin-Cannici, C. (1995). The Personality Assessment Inventory (PAI) and the detection of defensiveness. *Assessment, 2(4),* 333-342.

Conger, A.J., & Conger, J.C. (1996). Did too, did not!: Controversies in the construct validation of the PAI. *Journal of Psychopathology and Behavioural Assessment, 18,* 205-212.

Cortina, J.M., Doherty, M.L., Schmitt, N., Kaufman, G., & Smith, R.G. (1992). The "Big Five" personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology, 45,* 119-140.

Costa, P.T. & McCrae, R.R. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51(6),* 882-888.

Costa, P.T. & McCrae, R.R. (1992a). *NEO-PI-R professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P.T. & McCrae, R.R. (1992b). Normal personality assessment in clinical practice; The NEO Personality Inventory. *Psychological Assessment, 4(1),* 5-13.

Costa, P.T., & McCrae, R.R. (1995). Persons, places, and personality: Career assessment using the revised NEO Personality Inventory. *Journal of Career Assessment, 3(2),* 123-129.

Costello, R.M., Schneider, S.L., & Schoenfeld, L.S. (1993). Applicants fraud in law enforcement. *Psychological Reports, 73,* 179-183.

Coutts, L.M. (1990). Police hiring and promotion: Methods and outcomes. *Canadian Police College Journal, 14(2),* 98-122.

DuBois, C.L.Z., Sackett, P.R., Zedeck, S., & Fogli, L. (1993). Further exploration of typical and maximum performance criteria: Definitional issues, prediction, and white-black differences. *Journal of Applied Psychology, 78,* 205-211.

Durie, M. (1985). *Counselling Maori people.* Transcript of a presentation to the New Zealand Counselling and Guidance Association Hui.

Durie, M. (1994). *Whaiora: Maori health development.* Auckland: Oxford University Press.

Elion, V.H., & Megargee, E.I. (1975). Validity of the MMPI Pd scale among black males. *Journal of Consulting and Clinical Psychology, 43,* 166-172.

Fals-Stewart, W. (1996). The ability of individuals with psychoactive substance use disorders to escape detection by the Personality Assessment Inventory. *Psychological Assessment, 8(1),* 60-68.

Fals-Stewart, W., & Lucente, S. (1997). Identifying positive dissimulation by substance-abusing individuals on the Personality Assessment Inventory: A cross-validation study. *Journal of Personality Assessment, 68(2),* 455-469.

Fyfe, J.F., Greene, J.R., Walsh, W.F., Wilson, O.W., & McLaren, R.C. (1997). *Police Administration* (5th ed.). Boston: McGraw-Hill.

Gaines, L.K., Kappeler, V.E., & Vaughn, J.B. (1997). *Policing in America.* (2nd ed.). Cincinnati, OH: Anderson Publishing Co.

Gatewood, R.D., & Field, H.S. (1998). *Human resource selection.* (4th ed.). Orlando, FL: The Dryden Press.

Green, S.B., & Kelley, C.K. (1988). Racial bias in prediction with the MMPI for a juvenile delinquent population. *Journal of Personality Assessment, 52,* 263-275.

Groth-Marnat, G. (1997). *Handbook of Psychological Assessment (3rd ed).* New York: John Wiley & Sons.

Hargrave, G.E., & Hiatt, D. (1987). Law enforcement selection with the Interview, MMPI, and CPI: A study of reliability and validity. *Journal of Police Science and Administration, 15(2),* 110-117.

Hart, P.M., Wearing, A.J., & Headey, B. (1995). Police stress and wellbeing: Integrating personality, coping and daily work experiences. *Journal of Occupational and Organisational Psychology, 68,* 133-156.

Hays, J.R. (1997). Notes on the concurrent validation of the Personality Assessment Inventory in law enforcement. *Psychological Reports, 81,* 244-246.

Helmes, E. (1993). A modern instrument for evaluating psychopathology-The Personality Assessment Inventory: Professional Manual. *Journal of Personality Assessment, 61(2),* 414-417.

Henderson, N.D. (1979). Criterion-related validity of personality and aptitude scales: A comparison of validation results under voluntary and actual test conditions. In C.D. Spielberger (Ed.), *Police selection and evaluation: Issues and techniques* (pp.179-195). Washington, DC: Hemisphere.

Hiatt, D., & Hargrave, G.E. (1988). Predicting job performance problems with psychological screening. *Journal of Police Science and Administration, 16(2),* 122-125.

Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51,* 469-477.

Hogg, A., & Wilson, C. (1995). *Is the psychological screening of police applicants a realistic goal? - The successes and failures of psychological screening.* Payneham: National Police Research Unit.

Hough, L.M., Eaton, N.K., Dunette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75(5),* 581-595.

Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 209-244.

Jackson, D.N. (1971). The dynamics of structured personality tests. *Psychological Review, 78(3),* 229-248.

Kaplan, H.I., & Sadock, B.J. *Synopsis of Psychiatry: Behavioral Sciences/clinical psychiatry.* (8th ed.). Baltimore, ML: Williams & Wilkins.

Kleinman, L..S., & Gordon, M.E. (1986). An examination of the relationship between police training academy performance and job performance. *Journal of Police Science and Administration, 14,* 293-299.

Kline, P. (1993). *The handbook of psychological testing.* London: Routledge.

Leonard, V.A., & More, H.W. (1993). *Police organisation and management (8th ed.).* New York: The Foundation Press.

Lorr, M., & Strack, S. (1994). Personality profiles of police candidates. *Journal of Clinical Psychology, 50(2),* 200-207.

Maxwell, S.E., & Arvey, R.D. (1993). The search for predictors with high validity and low adverse impact: Compatible or incompatible goals. *Journal of Applied Psychology, 78,* 433-437.

McFarlane-Nathan, G. (1999). *FreMO: Framework for reducing Maori offending.* Wellington, New Zealand: Department of Corrections.

McQuilkin, J.I., Russell, V.L., Frost, A.G., & Faust, W.R. (1990). Psychological test validity for selecting law enforcement officers. *Journal of Police Science and Administration, 17(4),* 289-294.

Meier, R.D., Farmer, R.E., & Maxwell, D. (1987). Psychological screening of police candidates: Current perspectives. *Journal of Police Science and Administration, 15(3),* 210-215.

Millon, T. (1994). *Millon Clinical Multiaxial Inventory III: Manual for the MCMI-III.* Minneapolis, MN: National Computer Systems.

Morey, L.C. (1991). *The Personality Assessment Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L.C. (1995). Critical issues in construct validation: Comment on Boyle and Lennon (1994). *Journal of Psychopathology and Behavioral Assessment, 17,* 393-401.

Morey, L.C. (1996). *An interpretive guide to the Personality Assessment Inventory.* Odessa, FL: Psychological Assessment Resources.

Morey, L.C., & Lanier, V.W. (1998). Operating characteristics of six response distortion indicators for the Personality Assessment Inventory. *Assessment, 5(3),* 203-214.

Narrol, H.G., & Levitt, E.E. (1963). Formal assessment procedures in police selection. *Psychological Reports, 12,* 691-694.

National Health Committee. (1996). *Guidelines for the treatment and management of depression by primary healthcare professionals.* Wellington, New Zealand: National Health Committee.

New Zealand Police, (1998). *Attrition in the New Zealand police: Proceedings of the Disengagement Summit.* Wellington: New Zealand Police.

Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81(6),* 660-679.

Paulhus, D.L., & Reid, D.B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60,* 307-317.

Peebles, J., & Moore, R.J. (1998). Detecting socially desirable responding with the Personality Assessment Inventory: The Positive Impression Management scale and the Defensiveness Index. *Journal of Clinical Psychology, 54,* 621-628.

Piotrowski, C., & Belter, R.W. (1999). Internship training in psychological assessment: Has managed care had an impact? *Assessment, 6,* 381-389.

Piotrowski, C. (2000). How popular is the Personality Assessment Inventory in practice and training? *Psychological Reports, 86,* 65-66.

Poland, J.M. (1978). Police selection methods and the prediction of police performance. *Journal of Police Science and Administration, 6(4),* 374-393.

Police Education Advisory Council. (1998). *Police for the future: Review of recruitment and selection for the Queensland Police Service.* Brisbane: Criminal Justice Commission.

Retzlaf, P., & Gilbertini, M. (1987). Factor structure of the basic personality scales and common-item artifact. *Journal of Personality Assessment, 51,* 588-594.

Roberts, M.D., Thompson, J.A., & Johnson, M. (1999). *PAI Law enforcement, corrections, and public safety selection report module manual.* Odessa, Fl: Psychological Assessment Resources, Inc.

Robinson, H.M., Sigman, M.R., & Wilson, J.P. (1997). Duty-related stressors and PTSD symptoms in suburban police officers. *Psychological Reports, 81,* 835-845.

Rogers, R., Ornduff, S.R., & Sewell, K.W. (1993). Feigning specific disorders: A study of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment, 67,* 554-560.

Rogers, R., Sewell, K.W., Morey, L.C. & Ustad, K.L. (1997). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment, 67,* 629-640.

Rogers, R., Ustad, K.L., & Salekin, R.T. (1998). Convergent validity of the Personality Assessment Inventory: A study of emergency referrals in a correctional setting. *Assessment, 5,* 3-12.

Ross, S.R., Bailley, S.E., & Millis, S.R. (1997). Positive self-presentation effects and the detection of defensiveness on the NEO-PI-R. *Assessment, 4(4),* 395-408.

Rosse, J.G., Stecher, M.D., Miller, J.L., & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83(40),* 634-644.

Rothstein, H. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75,* 322-327.

Sarchione, C.D., Cuttler, M.J., Muchinsky, P.M., & Nelson-Gray, R.O. (1998). Prediction of dysfunctional job behaviors among law enforcement officers. *Journal of Applied Psychology, 83(6),* 904-912.

Saxe. S.J. & Reiser, M. (1976). A comparison of three police applicant groups using the MMPI. *Journal of Police Science and Administration, 4,* 419-425.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8(4),* 350-353.

Shusman. E.J., Inwald, R.E., & Knatz, H.F. (1987). A cross-validation study of police recruit performance as predicted by the IPI and MMPI. *Journal of Police Science and Adminstration, 15(2),* 162-169.

Smith, R.L., Ager, J.W., & Williams, D.L. (1992). Suppressor variables in multiple regression/correlation. *Educational and Psychological Measurement, 52(1),* 17-29.

Tabachnik, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*. (2nd ed.). New York: Harper & Row.

Topping, G.D., & O'Gorman, J.G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences, 23(1),* 117-124.

Violanti, J.M. & Aron, F. (1993). Sources of police stressors, job attitudes, and psychological distress. *Psychological Reports, 72,* 899-904.

Wang, E.W., Rogers, R., Giles, C.L., Diamond, P.M., Herrington-Wang, L.E., & Taylor, E.R. (1997). A pilot study of the Personality Assessment Inventory (PAI) in corrections: Assessment of malingering, suicide risk, and aggression in male inmates. *Behavioral Sciences and the Law, 15,* 469-482.

Watkins,C.E., Campbell, V.I., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 24,* 54-60.

White, L.J. (1996). Review of the Personality Assessment Inventory (PAI): A new psychological test for clinical and forensic assessment. *Australian Psychologist, 31(1),* 38-39.

Appendix A

Table 25. *PAI Full Scale Raw Scores for the Police Recruit Sample (N = 127).*

Scale	Mean	*SD*
Inconsistency (ICN)	4.36	2.80
Infrequency (INF)	3.25	2.11
Negative Impression (NIM)	.74	1.24
Positive Impression (PIM)	18.02	3.44
Somatic Complaints (SOM)	4.36	3.73
Anxiety (ANX)	4.36	6.65
Anxiety-Related Disorders (ARD)	15.03	5.57
Depression (DEP)	8.87	4.89
Mania (MAN)	22.57	7.65
Paranoia (PAR)	15.76	5.46
Schizophrenia (SCZ)	11.00	4.92
Borderline Features (BOR)	14.09	6.82
Antisocial Features (ANT)	16.60	7.92
Alcohol Problems (ALC)	5.44	3.88
Drug Problems (DRG)	2.97	2.96
Aggression (AGG)	11.73	5.38
Suicidal Ideation (SUI)	1.51	2.19
Stress (STR)	4.65	2.62
Nonsupport (NON)	2.90	2.71
Treatment Rejection (RXR)	16.98	3.09
Dominance (DOM)	20.85	3.87
Warmth (WRM)	25.04	4.08

Table 26. *PAI Subscale Raw Scores for the Police Recruit Sample (N = 127).*

Scale	Mean	*SD*
Somatic Complaints (SOM)		
Conversion (SOM-C)	.80	1.79
Somatization (SOM-S)	1.53	1.51
Health Concerns (SOM-H)	2.04	1.70
Anxiety		
Cognitive (ANX-C)	4.46	2.83
Affective (ANX-A)	4.81	2.87
Physiological (ANX-P)	3.35	2.00
Anxiety-Related Disorders (ARD)		
Obsessive-Compulsive (ARD-O)	8.39	2.98
Phobias (ARD-P)	4.58	2.26
Traumatic Stress (ARD-T)	2.06	2.42
Depression (DEP)		
Cognitive (DEP-C)	3.04	2.12
Affective (DEP-A)	2.04	1.77
Physiological (DEP-P)	3.79	2.77
Mania (MAN)		
Activity Level (MAN-A)	7.06	2.77
Grandiosity (MAN-G)	8.95	3.91
Irritability (MAN-I)	6.56	3.37
Paranoia (PAR)		
Hypervigilance (PAR-H)	7.16	2.60
Persecution (PAR-P)	2.39	1.82
Resentment (PAR-R)	6.20	2.50
Schizophrenia (SCZ)		
Psychotic Experiences (SCZ-P)	3.50	2.13
Social Detachment (SCZ-S)	4.25	2.34
Thought Disorder (SCZ-T)	3.24	2.32
Borderline Features (BOR)		
Affective Instability (BOR-A)	2.84	2.00
Identity Problems (BOR-I)	4.08	2.38
Negative Relationships (BOR-N)	3.91	2.56
Self-Harm (BOR-S)	3.26	2.17
Antisocial Features (ANT)		
Antisocial Behaviours (ANT-A)	4.88	3.67
Egocentricity (ANT-E)	4.20	2.39
Stimulus-Seeking (ANT-S)	7.51	3.86
Aggression (AGG)		
Aggressive Attitude (AGG-A)	2.95	2.70
Verbal Aggression (AGG-V)	7.48	2.72
Physical Aggression (AGG-P)	1.30	1.72

Table 27. *Intercorrelation Matrix for PAI Full Scales for the Police Recruit Sample (N = 127).*

	INF	NIM	PIM	SOM	ANX	ARD	DEP	MAN	PAR	SCZ	BOR	ANT	ALC	DRG	AGG	SUI	STR	NON	RXR	DOM	WRM
ICN	.03	.03	-.09	.06	.28**	.17*	.31**	.05	.08	.13	.14	-.02	.06	.69**	-.13	.26**	.06	.05	-.20**	-.19*	-.08
INF		-.06	.12	.06	.09	-.02	.15	-.07	.08	.01	.09	-.10	-.13	.07	-.06	.04	.09	.10	.01	-.02	.11
NIM			-.34**	.48**	.26**	.42**	.36**	.30**	.35**	.50**	.54**	.40**	.30**	.06	.28**	.34**	.34**	.36**	-.34**	-.14	-.07
PIM				-.30**	-.47**	-.53**	-.31**	-.48**	-.34**	-.40**	-.69**	-.49**	-.36**	-.03	-.48**	-.25**	-.30**	-.24**	.47**	-.06	.18*
SOM					.39**	.30**	.46**	.16	.24**	.34**	.43**	.28**	.34**	.18*	.16	.37**	.36**	.38**	-.43**	-.22*	.03
ANX						.63**	.67**	.20*	.44**	.54**	.58**	.14	.21*	.19*	.19*	.41**	.25**	.31**	-.59**	-.27**	-.26**
ARD							.52**	.36**	.44**	.58**	.63**	.28**	.30**	.16	.22*	.46**	.31**	.38**	-.54**	-.16	-.26*
DEP								.04	.44**	.48**	.61**	.14	.11	.22*	.16	.40**	.38**	.41**	-.53**	-.36**	-.22*
MAN									.22*	.28**	.49**	.55**	.34**	-.06	.38**	.24**	.19*	.06	-.19*	.42**	.09
PAR										.37**	.59**	.31**	.07**	.10	.32**	.18*	.30**	.36**	-.37**	-.07	-.33**
SCZ											.57**	.34**	.23*	.12	.30**	.32**	.30**	.41**	-.48**	-.26**	-.38**
BOR												.52**	.31**	.14	.43**	.43**	.52**	.48**	-.59**	.01	-.18*
ANT													.46**	.08	.35**	.16	.26**	.32**	-.23**	.23*	-.10
ALC														.07	.24**	.24**	.24**	.19*	-.26**	.09	-.05
DRG															-.02	.13	.06	.09	-.16	-.14	-.10
AGG																.11	.16	.13	-.17	.24**	-.19*
SUI																	.22**	.28**	-.39**	-.11	.03
STR																		.37**	-.35**	-.18*	-.12
NON																			-.42**	-.17	-.35*
RXR																				.21*	.15
DOM																					.20
WRM																					

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

INF Infrequency *ICN* Inconsistency *NIM* Negative Impression Management *PIM* Positive Impression Management *SOM* Somatic Complaints *ANX* Anxiety *ARD* Anxiety-Related Disorders *DEP* Depression *MAN* Mania *PAR* Paranoia *SCZ* Schizophrenia *BOR* Borderline Features *ANT* Antisocial Features *AGG* Aggression *SUI* Suicidal Ideation *STR* Stress *NON* Non-Support *RXR* Treatment Rejection *DOM* Dominance *WRM* Warmth

Table 28. *Correlation Between PAI Full and Subscales for the Police Recruit Sample (N = 127).*

	ICN	INF	NIM	PIM	SOM	ANX	ARD	DEP	MAN	PAR	SCZ	BOR	ANT	ALC	DRG	AGG	SUI	STR	NON	RXR	DOM	WRM
SOM-C	.03	.02	.48**	-.19*	.75**	.25**	.21*	.39**	.10	.13	.31**	.33**	.23*	.29**	.09	.08	.31**	.32**	.42**	.37**	-.22*	.06
SOM-S	.06	.00	.41**	-.31**	.72**	.44**	.32**	.39**	.15	.23*	.33**	.39**	.15	.27**	.09	.18*	.26**	.26**	.22*	.40**	-.28**	-.12
SOM-H	.05	.12	.17	-.17	.76**	.20*	.16	.24**	.12	.19*	.11	.24**	.25*	.19*	.22*	.11	.25**	.22*	.20*	.21*	.00	.11
ANX-C	.23**	.03	.22*	-.50**	.35**	.90**	.62**	.62**	.17	.44**	.42**	.54**	.16	.20*	.17	.20*	.33**	.24**	.25**	.55**	-.24**	-.20*
ANX-A	.35**	.06	.18*	-.34**	.31**	.90**	.53**	.59**	.15	.36**	.45**	.46**	.06	.13	.29**	.18*	.38**	.19*	.24**	.54**	-.24**	-.27**
ANX-P	.10	.19*	.31**	-.35**	.39**	.76**	.49**	.53**	.23*	.36**	.55**	.51**	.20*	.20*	.01	.09	.35**	.24**	.39**	.42**	-.21*	-.20*
ARD-O	.08	-.05	.35**	-.27**	.15	.27**	.73**	.18*	.32**	.30**	.36**	.35**	.17*	.20*	.07	.17	.25**	.08	.13	.27**	.02	-.13
ARD-P	.18*	.02	.18*	-.48**	.33**	.60**	.68**	.49**	.13	.30**	.48**	.46**	.14	.14	.20*	.10	.25**	.26**	.34**	.48**	-.29**	-.29**
ARD-T	.12	-.01	.36**	-.43**	.21*	.55**	.76**	.50**	.31**	.36**	.43**	.58**	.30*	.31**	.07	.19*	.53**	.37**	.38**	.45**	-.12	-.16
DEP-C	.29**	.10	.25**	-.20*	.23*	.49**	.40**	.68**	-.17	.33**	.42**	.33**	-.04	-.05	.21*	.05	.31**	.18*	.24**	.40**	-.46**	-.22*
DEP-A	.23**	.09	.38**	-.25**	.37**	.46**	.42**	.70**	-.04	.33**	.40**	.49**	.09	.12	.10	.08	.45**	.35**	.48**	.42**	-.32**	-.23*
DEP-P	.18*	.13	.20*	-.23**	.40**	.52**	.32**	.79**	.20*	.33**	.26**	.49**	.22*	.16	.16	.18*	.19*	.29**	.24**	.37**	-.07	-.07
MAN-A	.06	.07	.38**	-.36**	.24**	.31**	.37**	.28**	.73**	.21*	.36**	.55**	.56**	.31**	.01	.22*	.26**	.22*	.21*	.24**	.28**	.03
MAN-G	-.07	-.03	.03	-.19*	.04	-.09	-.01	-.23**	.78**	-.04	-.03	.13	.37**	.17	-.10	.24**	.06	.00	.16	.05	.48**	.27*
MAN-I	.15	-.18*	.31**	-.58**	.11	.28**	.51**	.11	.77**	.36**	.36**	.51**	.37**	.31**	-.02	.40**	.24**	.23**	.15	.28**	.17	-.14
PAR-H	-.02	.03	.35**	-.33**	.21*	.34**	.38**	.32**	.24**	.79**	.35**	.48**	.24**	.06	-.03	.31**	.14	.29**	.29**	.30**	.02	-.25**
PAR-P	.13	.08	.39**	-.42**	.22*	.45**	.42**	.42**	.29**	.76**	.32**	.56**	.35**	.15	.12	.31**	.23**	.29**	.30**	.35**	-.10	-.21*
PAR-R	.10	.10	.13	-.20*	.15	.28**	.24**	.32**	.01	.80**	.20*	.36**	.15	-.01	.14	.14	.07	.15	.28**	.23**	-.08	-.32**
SCZ-P	.08	.11	.29**	-.19*	.29**	.32**	.40**	.22*	.38**	.16	.72**	.37**	.33**	.25**	.08	.20*	.30**	.22*	.19*	.26**	-.08	-.01
SCZ-S	.06	.02	.24**	-.23**	.10	.28**	.37**	.30**	-.02	.31**	.68**	.32**	.11	-.01	.03	.19*	.11	.21*	.51**	.30**	-.18*	-.63**
SCZ-T	.12	-.09	.54**	-.44**	.36**	.56**	.49**	.52**	.26**	.33**	.77**	.55**	.31**	.24**	.14	.25**	.29**	.22*	.19*	.49**	-.30**	-.17
BOR-A	.15	-.01	.53**	-.61**	.30**	.46**	.49**	.54**	.35**	.43**	.41**	.78**	.36**	.30**	.04	.39**	.37**	.41**	.40**	.48**	.01	-.22*
BOR-I	.25*	-.05	.47**	-.52**	.47**	.55**	.62**	.58**	.30**	.46**	.54**	.77**	.36**	.28**	.21*	.26**	.45**	.42**	.36**	.63**	-.16	-.10
BOR-N	.01	.11	.36**	-.51**	.21*	.44**	.54**	.45**	.35**	.56**	.47**	.81**	.33**	.18*	.02	.35**	.27**	.45**	.49**	.40**	.00	-.24**
BOR-S	.02	.21*	.26**	-.44**	.29**	.27**	.21*	.24**	.47**	.28**	.27**	.62**	.50**	.20*	.15	.30**	.19*	.25**	.18*	.24**	.20*	.05
ANT-A	-.03	-.12	.35**	-.47**	.37**	.21*	.29**	.16	.38**	.28**	.33**	.44**	.78**	.33**	.11	.25**	.16	.23**	.26**	.28**	.05	-.14
ANT-E	-.02	.00	.35**	-.36**	.14	.05	.22*	.02	.52**	.27**	.25**	.39**	.76**	.37**	.03	.33**	.03	.08	.19*	.13	.30**	-.02
ANT-S	-.01	-.08	.28**	-.35**	.14	.08	.19*	.13	.45**	.19*	.24**	.41**	.84**	.42**	.05	.27**	.14	.27**	.30**	.12	.22*	-.06
AGG-A	-.03	-.02	.41**	-.51**	.27**	.37**	.36**	.43**	.33**	.32**	.34**	.55**	.28**	.18*	-.02	.81**	.26**	.24**	.19*	.30**	.03	-.17
AGG-V	-.20*	.00	-.05	-.25**	-.12	-.04	-.03	-.22*	.23*	.17	.13	.13	.14	.08	-.08	.72**	-.05	-.02	-.02	.09	.35**	-.16
AGG-P	-.05	-.14	.31**	-.31**	.26**	.06	.17	.16	.33**	.22*	.19*	.30**	.44**	.33**	.05	.71**	.03	.14	.16	.21*	.13	-.10

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

KEY *INF* Infrequency *ICN* Inconsistency *NIM* Negative Impression Management *PIM* Positive Impression Management *SOM* Somatic Complaints *ANX* Anxiety *ARD* Anxiety-Related Disorders *DEP* Depression *MAN* Mania *PAR* Paranoia *SCZ* Schizophrenia *BOR* Borderline Features *ANT* Antisocial Features *AGG* Aggression *SUI* Suicidal Ideation *STR* Stress *NON* Non-Support *RXR* Treatment Rejection *DOM* Dominance *WRM* Warmth *SOM-C* Conversion *SOM-S* Somatization *SOM-H* Health Concerns *ANX-C* Cognitive *ANX-A* Affective *ANX-P* Physiological *ARD-O* Obsessive-Compulsive *ARD-P* Phobias *ARD-T* Traumatic Stress DEP-C Cognitive DEP-A Affective DEP-P Physiological MAN-A Activity Level *MAN-G* Grandiosity *MAN-I* Irritability *PAR-H* Hypervigilance *PAR-P* Persecution *PAR-R* Resentment *SCZ-P* Psychotic Experiences *SCZ-S* Social Detachment *SCZ-T* Thought Disorder *BOR-A* Affective Stability *BOR-I* Identity Problems *BOR-N* Negative Relationships *BOR-S* Self-Harm *ANT-A* Antisocial Behaviours *ANT-E* Egocentricity *ANT-S* Stimulus-Seeking *AGG-A* Aggressive Attitude *AGG-V* Verbal Aggression *AGG-P* Physical Aggression

Table 29. *Correlation Between PAI Full Scales and NEO-PI-R Facet Scales for the Police Recruit Sample (N = 106).*

	ICN	INF	NIM	PIM	SOM	ANX	ARD	DEP	MAN	PAR	SCZ	BOR	ANT	ALC	DRG	AGG	SUI	STR	NON	RXR	DOM	WRM
N1	.08	-.14	.23*	-.31**	.31**	.45**	.40**	.25**	.22*	.35**	.24*	.39**	.21*	.20*	.12	.17	.12	.15	.09	-.38**	-.12	-.18
N2	.00	-.14	.26**	-.41**	.29**	.37**	.29**	.34**	.22*	.42**	.22*	.42**	.26**	.16	.00	.45**	.11	.08	.07	-.28**	-.04	-.24*
N3	.15	-.10	.26**	-.46**	.34**	.48**	.47**	.39**	.15	.32**	.31**	.45**	.23*	.21*	.15	.22**	.25**	.14	.20*	-.47**	-.19*	-.24*
N4	.08	-.10	.10	-.32**	.05	.19	.25*	.05	.09	.20*	.18	.21*	.22*	.05	.13	.11	-.08	.06	.06	-.11	-.16	-.23*
N5	-.03	-.07	.25**	-.37**	.33**	.26**	.30**	.21*	.30**	.35**	.20*	.41**	.24*	.30**	.14	.36**	.11	.01	.07	-.26**	.09	-.08
N6	.09	-.13	.38**	-.41**	.25**	.34**	.25**	.25**	.00	.33**	.31**	.34**	.17	.10	.18	.22*	.08	.08	.02	-.27**	-.28**	-.33**
E1	.08	-.02	-.07	-.02	-.02	-.01	-.09	-.12	.14	-.20*	-.22*	-.07	-.03	.02	.02	-.04	.04	.05	-.17	.04	.13	.54**
E2	.08	.14	-.13	-.12	.03	.08	-.02	.06	.04	.06	-.33**	.07	.00	.06	-.02	-.08	.06	.00	-.12	-.02	-.06	.24*
E3	.08	.12	-.06	-.02	.00	-.09	-.13	-.09	.33**	-.03	-.12	.03	.12	.06	.05	.12	.01	-.09	-.07	-.04	.47**	.26**
E4	.06	.24*	-.06	.04	-.07	-.04	-.20*	-.03	.19	-.02	-.19	.09	.10	-.05	.03	.02	-.13	.17	.08	.07	.38**	.12
E5	-.10	.09	.10	-.25**	.06	.11	.01	.09	.34**	.25**	-.03	.27**	.33**	.14	-.12	.24*	-.02	.24*	.10	-.16	.12	.03
E6	.00	-.11	.03	-.11	-.08	-.11	-.04	-.15	.26**	-.08	-.14	.02	.01	.07	-.06	.08	-.02	.06	-.18	.09	.20*	.39**
O1	.08	.07	.19	-.16	.16	.18	.19*	.14	.18	.05	.22*	.26**	.19	.18	.02	-.02	.15	.08	.14	-.23*	.11	-.01
O2	.02	.18	.10	-.09	.08	.03	.07	.05	.20*	.03	.00	.14	.04	-.01	.04	.00	.04	.18	-.01	-.19*	-.05	.29**
O3	.02	.02	.12	-.16	.01	.19	.14	.08	.21*	.06	-.02	.22**	-.10	-.02	.02	.03	.07	.13	-.08	-.11	.01	.40**
O4	.07	.18	-.12	.10	-.15	-.03	-.15	-.05	.04	-.17	-.10	.02	.02	-.01	.04	-.12	-.04	-.06	.05	-.05	.12	.23*
O5	-.04	.21*	.05	-.03	-.10	.04	-.05	.05	.16	.03	-.02	.11	-.01	-.08	-.18	.00	-.01	.15	.09	-.10	.19*	.21**
O6	.05	.06	.27**	-.03	.11	.15	.12	.14	.11	.02	.16	.19	.00	.12	.03	.01	.16	.18	.10	-.16	.03	.14
A1	-.05	-.13	-.20*	.12	-.25**	-.26**	-.25*	-.35**	-.10	-.42**	-.16	-.30**	-.17	.05	-.09	-.12	-.18	-.15	-.29**	.21*	.01	.29**
A2	-.01	-.04	-.25*	.27**	-.22*	-.18	-.25**	-.22*	-.32**	-.29**	-.31**	-.35**	-.44**	-.16	-.04	-.16	-.07	-.02	-.32**	.11	-.10	.28**
A3	-.06	-.07	-.08	.14	-.11	-.08	-.06	-.18	.02	-.20*	-.23*	-.21*	-.19	-.03	-.14	-.23*	-.08	-.13	-.29**	.11	.03	.52**
A4	-.07	.05	-.25*	.29**	-.14	-.10	-.18	-.15	-.33**	-.32**	-.21*	-.35**	-.39**	-.23*	-.06	-.38**	-.09	-.02	-.17	.12	-.27**	.20*
A5	.07	-.02	-.06	.07	.06	.10	.04	.18	-.25*	-.03	-.03	-.01	-.18	-.03	.06	-.24*	.14	.20	.02	-.02	-.41**	.05
A6	.09	-.08	.02	.09	-.03	-.02	.04	.08	-.10	-.14	-.12	-.04	-.17	-.13	.16	-.07	.06	.03	-.22*	-.09	-.13	.35**
C1	-.10	-.05	-.20*	.26**	-.09	-.23*	-.09	-.24*	.11	-.24*	-.24*	-.27**	-.15	-.01	-.08	-.19*	-.04	-.03	-.18	.21*	.26**	.35**
C2	-.08	-.19*	.08	.06	.05	-.20*	.08	-.17	.10	-.11	-.12	-.09	.14	.09	-.02	-.04	-.13	.01	-.07	.21*	.22*	.15
C3	-.01	-.07	-.12	.31**	-.11	-.18	-.09	-.21*	-.08	-.39**	-.24*	-.35**	-.15	-.06	-.05	-.28**	-.14	-.10	-.14	.32**	.03	.28**
C4	-.05	.00	-.16	.13	-.25**	-.11	-.11	-.21*	.22*	-.19*	-.27**	-.13	-.08	-.10	-.09	-.06	-.21*	-.05	-.26**	.21*	.32**	.25*
C5	.10	.12	-.04	.23*	-.14	-.04	-.01	-.06	.06	-.12	-.16	-.10	-.11	-.09	.03	-.17	-.04	.05	-.10	.19	.11	.32**
C6	.01	.12	-.36**	.46**	-.20*	-.23*	-.29**	-.19	-.25**	-.35**	-.27**	-.47**	-.40**	-.27**	.01	-.31**	-.16	-.11	-.08	.32**	.02	.09

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Key N1 Anxiety N2 Angry Hostility N3 Depression N4 Self-consciousness N5 Impulsiveness N6 Vulnerability E1 Warmth E2 Gregariousness E3 Assertiveness E4 Activity E5 Excitement-seeking E6 Positive Emotions O1 Fantasy O2 Aesthetics O3 Feelings O4 Actions O5 Ideas O6 Values A1 Trust A2 Straightforwardness A3 Altruism A4 Compliance A5 Modesty A6 Tender-mindedness C1 Competence C2 Order C3 Dutifulness C4 Achievement Striving C5 Self-discipline C6 Deliberation *INF* Infrequency *ICN* Inconsistency *NIM* Negative Impression Management *PIM* Positive Impression Management *SOM* Somatic Complaints *ANX* Anxiety *ARD* Anxiety-Related Disorders *DEP* Depression *MAN* Mania *PAR* Paranoia *SCZ* Schizophrenia *BOR* Borderline Features *ANT* Antisocial Features *AGG* Aggression *SUI* Suicidal Ideation *STR* Stress *NON* Non-Support *RXR* Treatment Rejection *DOM* Dominance *WRM* Warmth

Table 30a. *Correlation Between PAI Subscales and NEO-PI-R Facet Scales for the Police Recruit Sample (N = 106).*

	N1	N2	N3	N4	N5	N6	E1	E2	E3	E4	E5	E6	O1	O2	O3	O4	O5	O6
SOM-C	.13	.20*	.30**	-.04	.19	.23*	-.08	-.07	-.02	-.01	-.06	-.04	.17	.11	.07	-.04	-.03	.10
SOM-S	.32**	.34**	.37**	.12	.34**	.30**	-.05	.05	-.07	-.22*	.10	-.17	.09	.05	-.03	-.17	-.15	.12
SOM-H	.23*	.11	.13	.03	.22*	.05	.07	.08	.07	.07	.08	.01	.11	.04	.00	-.11	-.04	.03
ANX-C	.51**	.40**	.52**	.28**	.27**	.37**	.01	.12	-.13	-.06	.06	-.05	.16	-.03	.21*	-.14	-.01	.08
ANX-A	.36**	.32**	.42**	.10	.20*	.31**	-.03	.01	-.05	-.02	.09	-.13	.16	.05	.15	.12	.05	.16
ANX-P	.28**	.22*	.26**	.08	.23*	.15	-.01	.07	-.04	.00	.15	-.13	.11	.07	.12	-.05	.09	.17
ARD-O	.32**	.18	.24*	.12	.16	.13	-.08	-.04	.04	-.14	.11	-.01	.05	.09	.09	-.16	.03	-.03
ARD-P	.27**	.21*	.32**	.31**	.21*	.26**	-.08	.00	-.22*	-.14	-.11	-.08	.18	-.08	.07	-.17	-.16	.12
ARD-T	.24*	.22*	.44**	.12	.26**	.15	-.02	-.01	-.15	-.14	.00	.01	.20*	.13	.14	.02	.00	.19
DEP-C	.10	.19	.32**	.09	.10	.33**	-.06	.01	-.21*	-.17	-.04	-.21*	.07	.04	-.07	-.01	-.10	.06
DEP-A	.19	.23*	.29**	.02	.07	.22*	-.17	-.01	-.16	-.02	-.03	-.16	.14	-.06	.05	-.13	.02	.14
DEP-P	.26**	.31**	.25**	.00	.25*	.05	-.06	.11	.12	.09	.21*	.01	.10	.08	.16	.01	.15	.11
MAN-A	.20*	.10	.11	.01	.27**	-.03	.08	.12	.25**	.21*	.39**	.11	.22*	.13	.09	.06	.09	.12
MAN-G	-.02	.04	-.01	.00	.13	-.17	.21*	.06	.37**	.13	.19	.25*	.07	.16	.15	.09	.22*	.03
MAN-I	.32**	.35**	.24*	.20*	.28**	.22*	-.01	-.07	.12	.10	.23*	.19*	.12	.12	.21*	-.07	.03	.10
PAR-H	.35**	.35**	.28**	.21*	.29**	.22*	-.20*	-.09	-.11	-.04	.17	-.03	.06	.03	.14	-.19	.10	.05
PAR-P	.23*	.36**	.30**	.18	.32**	.24*	-.01	.20*	.15	.04	.26**	-.06	-.01	.10	.14	-.05	.10	.09
PAR-R	.22*	.26**	.17	.07	.23*	.28**	-.21*	.08	-.05	-.02	.17	-.10	.04	-.05	-.12	-.16	-.10	-.05
SCZ-P	.08	.05	.02	-.01	.18	-.03	.07	-.12	.09	-.07	.09	.09	.20*	.05	.02	.06	.03	.09
SCZ-S	.06	.16	.19	.19	-.08	.24*	-.42**	-.43**	-.28**	-.07	-.15	-.32**	.02	-.08	-.14	-.13	.06	.09
SCZ-T	.38**	.25**	.42**	.21*	.32**	.42**	-.09	-.13	-.06	-.23*	.00	-.06	.24*	.02	.08	-.12	-.14	.13
BOR-A	.27**	.46**	.37**	.13	.24*	.33**	-.14	.00	-.08	.04	.17	-.07	.16	.11	.14	.02	.18	.16
BOR-I	.42**	.34**	.47**	.23*	.30**	.39**	-.12	.03	-.06	-.07	.02	.00	.28**	.02	.13	-.05	-.06	.18
BOR-N	.34**	.39**	.40**	.23*	.35**	.23*	-.15	.01	.01	.01	.21*	.01	.13	.11	.22*	.01	.09	.09
BOR-S	.11	.07	.09	.04	.32**	.06	.20*	.17	.20*	.29**	.40**	.10	.22*	.20*	.15	.09	.13	.14
ANT-A	.30**	.39**	.31**	.19	.29**	.28**	-.07	.00	.08	-.07	.13	-.11	.13	.05	-.09	-.10	-.11	-.10
ANT-E	.15	.13	.12	.26**	.21	.14	-.06	.08	.14	.14	.27**	.13	.14	-.06	-.05	.00	-.01	.03
ANT-S	.05	.09	.11	.13	.08	.01	.03	-.05	.08	.18	.38**	.04	.17	.07	-.09	.13	.09	.07
AGG-A	.21*	.47**	.28**	.17	.32**	.24*	-.03	.00	.09	-.11	.17	-.01	.03	.12	.10	-.16	.13	.10
AGG-V	.02	.22*	.09	.05	.25**	.14	-.05	-.08	.10	.11	.16	.12	-.07	-.16	-.01	-.04	-.12	-.15
AGG-P	.17	.36**	.12	.07	.27**	.10	-.05	-.09	.05	.01	.25**	.06	-.02	.05	-.09	-.10	-.01	.10

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

KEY N1 Anxiety N2 Angry Hostility N3 Depression N4 Self-consciousness N5 Impulsiveness N6 Vulnerability E1 Warmth E2 Gregariousness E3 Assertiveness E4 Activity E5 Excitement-seeking E6 Positive Emotions O1 Fantasy O2 Aesthetics O3 Feelings O4 Actions O5 Ideas O6 Values A1 Trust A2 Straightforwardness A3 Altruism A4 Compliance A5 Modesty A6 Tender-mindedness C1 Competence C2 Order C3 Dutifulness C4 Achievement Striving C5 Self-discipline C6 Deliberation *SOM-C* Conversion *SOM-S* Somatization *SOM-H* Health Concerns *ANX-C* Cognitive *ANX-A* Affective *ANX-P* Physiological *ARD-O* Obsessive-Compulsive *ARD-P* Phobias *ARD-T* Traumatic Stress DEP-C Cognitive DEP-A Affective DEP-P Physiological MAN-A Activity Level *MAN-G* Grandiosity *MAN-I* Irritability *PAR-H* Hypervigilance *PAR-P* Persecution *PAR-R* Resentment *SCZ-P* Psychotic Experiences *SCZ-S* Social Detachment *SCZ-T* Thought Disorder *BOR-A* Affective Stability *BOR-I* Identity Problems *BOR-N* Negative Relationships *BOR-S* Self-Harm *ANT-A* Antisocial Behaviours *ANT-E* Egocentricity *ANT-S* Stimulus-Seeking *AGG-A* Aggressive Attitude *AGG-V* Verbal Aggression *AGG-P* Physical Aggression

Table 30b. *Correlation Between PAI Subscales and NEO-PI-R Facet Scales for the Police Recruit Sample (N = 106).*

	A1	A2	A3	A4	A5	A6	C1	C2	C3	C4	C5	C6
SOM-C	-.19	-.27**	-.03	-.13	.05	-.02	-.14	.05	-.08	-.20*	-.08	-.20*
SOM-S	-.21*	-.13	-.13	-.08	.03	-.01	-.15	.03	-.20*	-.25**	-.20*	-.19
SOM-H	-.18	-.13	-.07	-.12	.05	-.05	.04	.05	.02	-.13	-.03	-.06
ANX-C	-.27**	-.12	.00	-.08	.05	.02	-.24*	-.15	-.17	-.02	-.07	-.24*
ANX-A	-.24*	-.18	-.11	-.07	.08	.04	-.19	-.17	-.15	-.11	-.02	-.20*
ANX-P	-.17	-.20*	-.11	-.10	.12	-.15	-.15	-.15	-.13	-.17	.00	-.15
ARD-O	-.21*	-.18	.04	-.10	-.14	.02	.11	.29**	.06	.06	.14	-.13
ARD-P	-.07	-.17	-.17	-.17	.11	-.10	-.24*	-.15	-.22*	-.23*	-.17	-.20*
ARD-T	-.21*	-.18	.00	-.10	.15	.15	-.09	-.01	-.07	-.11	-.03	-.29**
DEP-C	-.27**	-.03	-.16	-.02	.27**	.09	-.30**	-.18	-.15	-.30**	-.12	-.12
DEP-A	-.24*	-.13	-.17	-.02	.16	.06	-.17	-.14	-.14	-.15	.04	-.11
DEP-P	-.28**	-.31**	-.08	-.21*	-.01	.02	-.07	-.08	-.17	-.04	-.03	-.17
MAN-A	-.15	-.30**	-.02	-.24*	-.15	-.12	-.03	.04	-.01	.09	.06	-.27**
MAN-G	.01	-.12	.20*	-.15	-.22*	.04	.29**	.10	.01	.29**	.08	-.03
MAN-I	-.11	-.31**	-.16	-.34**	-.17	-.16	-.04	.06	-.18	.09	-.01	-.30**
PAR-H	-.35**	-.26**	-.10	-.28**	.02	-.07	-.13	-.12	-.21*	-.11	-.09	-.29**
PAR-P	-.29**	-.31**	-.16	-.31**	-.11	-.14	-.14	-.09	-.36**	-.09	-.05	-.35**
PAR-R	-.33**	-.14	-.21*	-.17	.00	-.12	-.26**	-.06	-.33**	-.23*	-.11	-.19*
SCZ-P	.06	-.17	.03	-.13	-.07	-.02	-.02	-.06	-.07	-.12	-.07	-.21*
SCZ-S	-.18	-.19	-.43**	-.15	-.07	-.24	-.22*	-.10	-.19*	-.23*	-.09	.01
SCZ-T	-.20*	-.30**	-.08	-.17	.07	.02	-.26**	-.10	-.23*	-.21*	-.18	-.38**
BOR-A	-.20*	-.23*	-.23*	-.32**	-.04	.03	-.25*	-.05	-.24*	-.13	-.05	-.32**
BOR-I	-.25*	-.23*	-.10	-.20*	.02	.08	-.27**	-.05	-.32**	-.15	-.07	-.38**
BOR-N	-.37**	-.37**	-.24*	-.26**	-.06	-.12	-.24*	-.06	-.32**	-.09	-.08	-.35**
BOR-S	-.06	-.21*	-.07	-.26**	.06	-.08	-.05	-.10	-.16	-.01	-.09	-.36**
ANT-A	-.26**	-.39**	-.21*	-.33**	-.14	-.09	-.17	.06	-.24*	-.15	-.26**	-.37**
ANT-E	-.07	-.38**	-.13	-.40**	-.23*	-.19	-.12	.16	-.13	.05	.00	-.30**
ANT-S	-.06	-.30**	-.10	-.24*	-.09	-.14	-.06	.13	.01	-.06	.03	-.27**
AGG-A	-.19*	-.12	-.22*	-.28**	-.13	.00	-.16	-.07	-.25**	-.07	-.20*	-.26**
AGG-V	.00	-.08	-.21*	-.29**	-.26**	-.12	-.19	-.06	-.29**	-.06	-.09	-.26**
AGG-P	-.10	-.21*	-.11	-.34**	-.17	-.04	-.11	.06	-.08	-.03	-.12	-.18

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Key N1 Anxiety N2 Angry Hostility N3 Depression N4 Self-consciousness N5 Impulsiveness N6 Vulnerability E1 Warmth E2 Gregariousness E3 Assertiveness E4 Activity E5 Excitement-seeking E6 Positive Emotions O1 Fantasy O2 Aesthetics O3 Feelings O4 Actions O5 Ideas O6 Values A1 Trust A2 Straightforwardness A3 Altruism A4 Compliance A5 Modesty A6 Tender-mindedness C1 Competence C2 Order C3 Dutifulness C4 Achievement Striving C5 Self-discipline C6 Deliberation *SOM-C* Conversion *SOM-S* Somatization *SOM-H* Health Concerns *ANX-C* Cognitive *ANX-A* Affective *ANX-P* Physiological *ARD-O* Obsessive-Compulsive *ARD-P* Phobias *ARD-T* Traumatic Stress DEP-C Cognitive DEP-A Affective DEP-P Physiological MAN-A Activity Level *MAN-G* Grandiosity *MAN-I* Irritability *PAR-H* Hypervigilance *PAR-P* Persecution *PAR-R* Resentment *SCZ-P* Psychotic Experiences *SCZ-S* Social Detachment *SCZ-T* Thought Disorder *BOR-A* Affective Stability *BOR-I* Identity Problems *BOR-N* Negative Relationships *BOR-S* Self-Harm *ANT-A* Antisocial Behaviours *ANT-E* Egocentricity *ANT-S* Stimulus-Seeking *AGG-A* Aggressive Attitude *AGG-V* Verbal Aggression *AGG-P* Physical Aggression

Table 31. *Correlation Between NEO-PI-R Domain Scales and PAI Subscales for the Police Recruit Sample (N = 106).*

	N	E	O	A	C
SOM-C	.23*	-.08	.10	-.16	-.15
SOM-S	.41**	-.10	-.03	-.14	-.22*
SOM-H	.18	.10	.00	-.13	-.03
ANX-C	.52**	-.03	.07	-.11	-.21*
ANX-A	.38**	-.05	.18	-.13	-.19*
ANX-P	.27**	.00	.14	-.16	-.17
ARD-O	.25**	-.03	.02	-.14	.12
ARD-P	.35**	-.18	-.03	-.15	-.27**
ARD-T	.32**	-.10	.18	-.06	-.14
DEP-C	.25*	-.20*	-.02	-.03	-.27
DEP-A	.22*	-.16	.04	-.11	-.16
DEP-P	.26**	.13	.17	-.22*	-.13
MAN-A	.16	.31**	.19	-.25*	-.03
MAN-G	.00	.34**	.22	-.06	.16
MAN-I	.35**	.15	.14	-.32**	-.09
PAR-H	.38**	-.08	.06	-.25**	-.22*
PAR-P	.36**	.17	.11	-.33**	-.26**
PAR-R	.27**	-.04	-.13	-.24*	-.27**
SCZ-P	.07	05	.12	-.09	-.12
SCZ-S	.15	-.46**	-.07	-.31**	-.18
SCZ-T	.44**	-.16	.04	-.18	-.31**
BOR-A	.40**	-.03	.21*	-.26**	-.24*
BOR-I	.47**	-.06	.12	-.19	-.29**
BOR-N	.43**	.03	.18	-.36**	-.27**
BOR-S	.17	.36**	.26**	-.16	-.18
ANT-A	.39**	-.02	-.06	-.36**	-.26**
ANT-E	.23*	.20*	.00	-.36**	-.08
ANT-S	.11	.17	.11	-.23*	-.06
AGG-A	.39**	.04	.08	-.24*	-.23*
AGG-V	.18	.11	-.14	-.23*	-.22*
AGG-P	.25**	.07	-.03	-.25**	-.11

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

KEY *SOM-C* Conversion *SOM-S* Somatization *SOM-H* Health Concerns *ANX-C* Cognitive *ANX-A* Affective *ANX-P* Physiological *ARD-O* Obsessive-Compulsive *ARD-P* Phobias *ARD-T* Traumatic Stress DEP-C Cognitive DEP-A Affective DEP-P Physiological MAN-A Activity Level *MAN-G* Grandiosity *MAN-I* Irritability *PAR-H* Hypervigilance *PAR-P* Persecution *PAR-R* Resentment *SCZ-P* Psychotic Experiences *SCZ-S* Social Detachment *SCZ-T* Thought Disorder *BOR-A* Affective Stability *BOR-I* Identity Problems *BOR-N* Negative Relationships *BOR-S* Self-Harm *ANT-A* Antisocial Behaviours *ANT-E* Egocentricity *ANT-S* Stimulus-Seeking *AGG-A* Aggressive Attitude *AGG-V* Verbal Aggression *AGG-P* Physical Aggression *N* Neuroticism *E* Extraversion *O* Openness to Experience *A* Agreeableness *C* Conscientiousness

Table 32. *Correlation Between PAI Full Scales and NEO-PI-R Domain Scales for the Police Recruit Sample (N = 106).*

	N	E	O	A	C
ICN	.08	.05	.05	-.02	-.03
INF	-.15	.11	.20*	-.07	-.02
NIM	.33**	-.05	.16	-.22*	-.19
PIM	-.51**	-.13	-.10	.25**	.33**
SOM	.36**	-.02	.02	-.19	-.17
ANX	.46**	-.04	.15	-.14	-.23*
ARD	.43**	-.13	.08	-.17	-.11
DEP	.34**	-.08	.11	-.17	-.25*
MAN	.23*	.36**	.25**	-.27**	.03
PAR	.44**	.00	.00	-.35**	-.32**
SCZ	.31**	-.28**	.04	-.27**	-.29**
BOR	.49**	.10	.25**	-.33**	-.33**
ANT	.30**	.14	.03	-.39**	-.18
ALC	.23*	.08	.05	-.15	-.09
DRG	.16	-.03	-.01	-.04	-.05
AGG	.35**	.10	-.03	-.30**	-.24*
SUI	.13	-.02	.09	-.06	-.17
STR	.11	.10	.19	-.02	-.06
NON	.12	-.11	.07	-.32**	-.19
RXR	-.40**	-.01	-.22*	.12	.33**
DOM	-.14	.35**	.11	-.21*	.22*
WRM	-.28**	.43**	.36**	.41**	.32**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

KEY *INF* Infrequency *ICN* Inconsistency *NIM* Negative Impression Management *PIM* Positive Impression Management *SOM* Somatic Complaints *ANX* Anxiety *ARD* Anxiety-Related Disorders *DEP* Depression *MAN* Mania *PAR* Paranoia *SCZ* Schizophrenia *BOR* Borderline Features *ANT* Antisocial Features *AGG* Aggression *N* Neuroticism *E* Extraversion *O* Openness to Experience *A* Agreeableness *C* Conscientiousness

Appendix B

Personality Test Research

Information Sheet

1. What is this study about and who is doing it?

This study is being undertaken as part of my Master of Arts degree in psychology. The study is intended to trial a personality test that may be used as an assessment measure for the New Zealand Police. This is not a test of intelligence, but rather assesses a range of responses found in the wider population.

My name is Nick Lascelles and I am a postgraduate student at Massey University training as a clinical psychologist. This research combines an interest in the police with an interest in the field of psychological testing. My supervisor is Ross Flett, Senior Lecturer at Massey University. You can contact by mail through the School of Psychology, Massey University, or you can phone me on (06) 359 0756 or Ross (06) 350 4127, or on email at nicklas@ihug.co.nz or at R.A.Flett@massey.ac.nz

This research is supported by Police Psychological Services on the basis that I can undertake this research as an independent researcher.

2. What will I be asked to do?

You will be given a psychological test that is intended to measure a range of psychological characteristics. This will take about an hour. The scores from the psychological test completed during your application and your results from the Royal New Zealand Police College will also be included in this research.

Your completion of the test will be taken as giving consent to participate in this study.

3. What rights do I have as a participant in this study?

- You have the right to decline to participate, refuse to answer particular questions or to withdraw from the study at any time
- You have the right to ask questions about the study at any time during participation
- You provide information on the understanding that it is confidential and to be used only by the researchers.
- You have the right to be given a summary of the research findings

4. What guarantees do I have regarding confidentiality?

All data gathered in the course of this study is confidential, adhering to the requirements of the Privacy Act 1993 and the Code of Ethics of the New Zealand Psychologists Board. The test is being given for research purposes only and the results will not be used by the Police for selection or assessment purposes. The Police will be not be able to find out the scores of any individual and only I will have access to the identity of anyone participating in this research.

When completing the test, you will be asked to put your name on the test sheet. However, these test sheets will not be viewed by anyone other than myself and no one else will be able to identify you including Police Psychological Services.

There is a very limited set of circumstances in which any participant's test scores and identity would be released to the Police. This is in the unlikely event that a participant's test scores suggested they were an imminent danger to themselves or another person. If a participant's scores raised serious concern, a senior psychologist at Massey University would be consulted before further action was taken.

5. What do I get from this research?

Once I have completed my research, you will be given a summary of the findings. If you request it, I will also provide you with an interpretation of your test results. There is a form provided at the back of this handout to request this.

Thank you for your participation,

Nick Lascelles

Personality Test Research

Request for Feedback

Please provide a mailing address for the period following completion of your training. This is so that I may send you a copy of the research findings and any other information you may request. If you do not wish to receive anything from me, leave this blank.

Name and Mailing Address: .....

.....

.....

.....

Request for additional information (please tick if required)

☐ I would like to be given an interpretation of my scores for the test I have completed. This is in addition to a summary of the research findings which all participants will receive.

Appendix C

Table 33. *Description of PAI Full Scales (Morey, 1996).*

Scale (designation)	Description
Validity Scales	
Inconsistency (ICN)	Assesses consistency of responding with pairs of highly correlated items.
Infrequency (INF)	Assesses for careless or random responding and includes items unrelated to psychopatholoy which have extremely high or low endorsement rates.
Negative Impression (NIM)	Assesses for exaggerated unfavourable presentation or malingering. Items have very low endorsement in clinical subjects.
Positive Impression (PIM)	Assesses for very favourable presentation or reluctance to admit minor flaws.
Clinical Scales	
Somatic Complaints (SOM)	Preoccupation with health matters and somatic complaints.
Anxiety (ANX)	Phenomenology and observable signs of anxiety.
Anxiety-Related Disorders (ARD)	Symptoms and behaviours of phobias, traumatic stress, and obsessive-compulsive disorders.
Depression (DEP)	Symptoms and phenomenology of depressive disorders.
Mania (MAN)	Affective, cognitive, and behavioural symptoms of mania and hypomania.
Paranoia (PAR)	Symptoms of paranoid disorders and paranoid personality characteristics.
Schizophrenia (SCZ)	Symptoms relating to broad range of schizophrenic disorders.
Borderline Features (BOR)	Characteristics indicating a borderline personality type.
Antisocial Features (ANT)	History of illegal acts and difficulties with authority, lack of empathy and loyalty, instability and excitement seeking.
Alcohol Problems (ALC)	Problematic consequences of alcohol use and aspects of alcohol dependence.
Drug Problems (DRG)	Problematic consequences of illegal and prescription drugs and aspects of drug dependence.
Treatment Scales	
Aggression (AGG)	Characteristics and attitudes related to anger, assertiveness, hostility, and aggression.
Suicidal Ideation (SUI)	Suicidal ideation including hopelessness and thoughts of suicide.
Stress (STR)	Impact of recent stressors in major life areas.
Nonsupport (NON)	Lack of perceived social support.
Treatment Rejection (RXR)	Attitudes and attributes theoretically predictive of interest and motivation in making changes of a psychological nature.
Interpersonal Scales	
Dominance (DOM)	Extent to which person is controlling and independent in personal relationships.
Warmth (WRM)	Extent to which a person is interested in supportive and empathic personal relationships.

Table 34a. *Description of PAI Subscales (Morey, 1996).*

Scale (designation)	Description
Somatic Complaints (SOM)	
Conversion (SOM-C)	Symptoms associated with conversion disorder.
Somatization (SOM-S)	Frequent occurrence of common physical symptoms and vague complaints of ill health and fatigue.
Health Concerns (SOM-H)	Preoccupation with health and physical problems.
Anxiety (ANX)	
Cognitive (ANX-C)	Ruminative worry and concern about current issues that result in impaired concentration and attention.
Affective (ANX-A)	Experience of tension, difficulty relaxing, and fatigue as result of high perceived stress.
Physiological (ANX-P)	Physical signs of tension and stress, such as sweaty palms, trembling hands, and shortness of breath.
Anxiety-Related Disorders (ARD)	
Obsessive-Compulsive (ARD-O)	Intrusive thoughts or behaviours, rigidity, indecisiveness, perfectionism, and affective constriction.
Phobias (ARD-P)	Common phobic fears.
Traumatic Stress (ARD-T)	Experience of past traumatic events that cause ongoing distress, and have caused lasting change or damage.
Depression (DEP)	
Cognitive (DEP-C)	Thoughts of worthlessness, hopelessness, and personal failure, as well as indecisiveness and difficulties in concentration.
Affective (DEP-A)	Feelings of sadness, loss of interest in normal activities, and anhedonia.
Physiological (DEP-P)	Physical functioning including appetite and sleep.
Mania (MAN)	
Activity Level (MAN-A)	Overinvolvement in wide range of activities in a disorganised manner, and experience of accelerated thought processes and behaviour.
Grandiosity (MAN-G)	Inflated self-esteem, expansiveness, and belief that one has special and unique skills or talents.
Irritability (MAN-I)	Strained relationships due to respondent's frustration with others inability or unwillingness to keep up with possibly unrealistic plans, demands or ideas.

Table 34b. *Description of PAI Subscales (Morey, 1996).*

Scale (designation)	Description
Paranoia (PAR)	
Hypervigilance (PAR-H)	Suspiciousness and the tendency to look for real or imagined slights by others.
Persecution (PAR-P)	Beliefs about being treated unfairly, and that there is a concerted attempt by others to undermine one's interests.
Resentment (PAR-R)	Bitterness and cynicism in interpersonal relationships, and a tendency to hold grudges and externalise blame.
Schizophrenia (SCZ)	
Psychotic Experiences (SCZ-P)	Experience of unusual perceptions and sensations, magical thinking, and/or any other unusual ideas that may involve delusional beliefs.
Social Detachment (SCZ-S)	Social isolation, discomfort, and awkwardness in social interactions.
Thought Disorder (SCZ-T)	Confusion, concentration problems, and disorganisation of thought processes.
Borderline Features (BOR)	
Affective Instability (BOR-A)	Emotional responsiveness, rapid mood changes, and poor emotional control.
Identity Problems (BOR-I)	Uncertainty about major life issues and feelings of emptiness, unfulfillment, and an absence of purpose.
Negative Relationships (BOR-N)	History of ambivalent, intense relationships in which one felt exploited or betrayed.
Self-Harm (BOR-S)	Impulsivity in areas that have high potential for negative consequences.
Antisocial Features (ANT)	
Antisocial Behaviours (ANT-A)	History of antisocial acts and involvement in illegal activities.
Egocentricity (ANT-E)	Lack of empathy or remorse and a generally exploitative approach to interpersonal relationships.
Stimulus-Seeking (ANT-S)	Craving for excitement and sensation, a low tolerance for boredom, and a tendency to be reckless and risk-taking.
Aggression (AGG)	
Aggressive Attitude (AGG-A)	Hostility, poor control over anger expression, and a belief in the instrumental use of aggression.
Verbal Aggression (AGG-V)	Verbal expressions of anger ranging from assertiveness to abusiveness, and a readiness to express anger to others.
Physical Aggression (AGG-P)	Tendency to physically display anger, including property damage, physical fights, and threats of violence.

Table 35a. *Descriptions of NEO-PI-R Domain and Facet Scales (Costa & McCrae, 1992).*

Scale (designation)	Description
Neuroticism (N)	Emotional adjustment vs maladjustment and emotional instabilty. Higher scores show a tendency to experience negative emotional states and may be less able to cope with stress. Low scores are usually calm, even-tempered and relaxed.
Anxiety (N1)	Apprehension, fearfulness, nervousness, tension, and proneness to worry.
Angry Hostility (N2)	Tendency to experience anger, frustration, and bitterness.
Depression (N3)	Tendency to experience depressive affect.
Self-consciousness (N4)	Predisposition to experience shame and embarrassment.
Impulsiveness (N5)	Inability to control cravings and urges.
Vulnerability (N6)	Vulnerability to stress.
Extraversion	Extraverts are sociable, assertive, active, talkative, optimistic and energetic. Introverts are not the opposite of an extraverts, but lack some extravert qualities. They may be reserved rather than unfriendly, independent rather than followers, and may prefer their own company rather than being shy.
Warmth (E1)	Warm people are affectionate and friendly while low scorers are more reserved and distant.
Gregariousness (E2)	Preference for the company of others. Low scorers tend to be loners.
Assertiveness (E3)	High scorers are dominant, forceful and assertive while low scorers prefer to stay in the background and let others do the talking.
Activity (E4)	A high activity score indicates a high level of energy and a fast paced life while low scorers are more leisurely and relaxed.
Excitement-Seeking (E5)	High scorers require excitement and stimulation.
Positive Emotions (E6)	Tendency to experience positive emotions such as joy, love, happiness and excitement. Low scorers are not necessarily unhappy, but merely less exuberant and high spirited.
Openness	Contains elements of imagination, creativity, sensitivity, intellectual curiosity, and independent thought. Low scorers tend to be conservative and conventional with a preference for the familiar rather than the novel.
Fantasy (O1)	High scorers have a vivid imagination and an active fantasy life.
Aesthetics (O2)	High scorers have a deep appreciation for art and beauty while low scorers are relatively insensitive in this area.
Feelings (O3)	High scores indicate receptivity to ones own inner feelings and emotions.
Actions (04)	High scorers prefer novelty and variety while low scorers dislike change and stick to the familiar.
Ideas (O5)	Intellectual curiosity, openmindedness and the willingness to consider new ideas. Low scorers have limited curiosity and if highly intelligent, focus on limited topics.
Values (O6)	Readiness to reexamine social, political, and religious values.

Table 35b. *Descriptions of NEO-PI-R Domain and Facet Scales (Costa & McCrae, 1992).*

Scale (designation)	Description
Agreeableness	Consists of interpersonal tendencies with a high scorer being altruistic, sympathetic to others and eager to help. The low scorer is egocentric, antagonistic, skeptical and competitive.
Trust (A1)	High scorers tend to believe that others are honest and well-intentioned while low scorers are more skeptical.
Straightforwardness (A2)	High scorers are frank and sincere while low scorers are more willing to manipulate others through flattery and deception.
Altruism (A3)	High scorers have an active concern for the welfare of others and are generous and willing to assist others. Low scorers are more self-centred and reluctant to get involved in the problems of others.
Compliance (A4)	Involves characteristic reactions to interpersonal conflict. High scorers tend to defer to others, inhibit aggression, and may be seen as meek and mild. Low scorers are aggressive and prefer to compete rather than cooperate.
Modesty (A5)	High scorers are humble and self-effacing while not lacking in confidence or self-esteem. Low scorers believe they are superior and may be considered arrogant or conceited.
Tender-Mindedness (A6)	Measures attitudes of sympathy and concern for others. Low scorers are more hardheaded and less moved by appeals for pity.
Conscientiousness	Conscientious people tend to be determined, reliable, punctual, and capable. High scorers are likely to be high achievers, but may also be fastidious, compulsively neat or a workaholic. Low scorers are not necessarily lazy, but take a more relaxed, lackadaisical approach to reaching their goals.
Competence (C1)	Measures the belief that one is capable, sensible, prudent, and effective.
Order (C2)	High scorers are neat, tidy and well organised while low scorers are disorganised and describe themselves as unmethodical.
Dutifulness (C3)	High scorers strictly adhere to their ethical principles and moral obligations. Low scorers may be more casual about such things and may be somewhat unreliable.
Achievement Striving (C4)	Measures aspirations, ambition, diligence and tendency to work hard. High scorers are diligent and purposeful and have a sense of direction in life. They may become workaholics however. Low scorers are lackadaisical and perhaps lazy, but usually content.
Self-Discipline (C5)	Measures the ability to carry a task through to completion despite boredom and other obstacles.
Deliberation (C6)	High scorers are cautious and deliberate while low scorers are often hasty and act without thinking.