

cerno



## Stage One Report

Advice on Police Standards and Assessment Practice

New Zealand Police

10.2007



# **POLICE STANDARDS AND ASSESSMENT PRACTICE (TO PROBATIONARY CONSTABLE LEVEL)**

## **Terms of Reference**

### **PURPOSE**

The Minister of Police has requested independent expert advice on the standards and assessment practices for Police applicants, Police recruits and probationary constables to determine whether they meet best practice.

### **Provide Advice**

The scope of the advice will include:

- The processes Police use to determine minimum standards for Police applicants, Police recruits and probationary constables.
- The processes Police use to determine assessment methods (where assessment methods means the tools and techniques used to assess whether police applicants, recruits, probationary constables meet the minimum standards).
- The procedures Police use to determine whether the application of assessment methods is consistently applied, nationally.

This work will take into account the results of an independent study on the efficacy of the data, methodology and conclusions in the reports by Wayne Annan, General Manager Human Resources, and Senior Sergeant Iain Saunders, RNZPC

### **Reporting**

- Police will be provided with the opportunity to comment on the draft report before Dr Dave George finalises and submits the report to the Minister of Police.
- The advice will be provided in a report to the Minister of Police, supported by a verbal briefing.

Timeframe: 30 September 2007.

### **Phase Two - Monthly Reporting**

Phase two will provide advice to the Minister of Police on the regular reporting regime being put in place by Police on applicant, recruit and probationary constable progress.

Timeframe: 12 months commencing 30 September 2007.

# 1. Executive Summary

This independent advice considers the standards and practices the New Zealand Police use to assess Recruits up to Probationary Constable level. The advice examines how Police have determined minimum standards, how they have chosen their assessment methods and how they ensure the assessment process is nationally consistent.

## **Determination of minimum standards**

The first question considered was whether “A systematic and robust process was used to identify the minimally acceptable knowledge, skills, attributes and other personal characteristics required of Police Recruits, Probationary Constables and Constables”.

It appears that Police have applied robust requirement identification processes to key elements of the policing role (for example, conducting criminal investigations, responding to traffic emergencies, conducting lawful interviews, and physical fitness). However, this appears to have been ad hoc with no evidence of any attempt to determine the full breadth and relative importance of the knowledge, skills and abilities required of Police Recruits.

While the major job requirements seem to have been covered, in absence of a systematic process it is not possible to state definitively that the minimally acceptable standards Police are using for assessment accurately reflect the complete requirements.

## **Assessment methods**

The second question considered was whether “A systematic and robust process was used to identify the assessment methods”.

While the range of assessment methods is broad, no evidence was provided of a systematic process being used to confirm that the range of assessment methods covers all of the necessary knowledge, skills and attributes. The assessment tools appear to have been chosen in response to particular needs, rather than as part of an overall integrated plan. In addition, processes to ensure these methods remain consistent with organisational requirements are not in place, risking a reduction in the relevance and effectiveness of assessment procedures over time.

In terms of the specific assessment methods, four of the fourteen assessment methods used to support the recruitment decision do not appear to meet baseline standards. The reliability and validity of these four tools is uncertain, and they lack soundly-based user guidelines and pass marks – severely limiting the usefulness of these specific tools. Of particular concern are the Behavioural Interview and the processes used for setting Reasoning Ability pass marks. These four tools may contribute to a higher than necessary variability in the quality of people entering the Police College – in terms of recruiting people who require more assistance than desirable and also missing out on potentially good Recruits.

However, all of the methods used post recruitment – the checks and balances that ensure the quality of people appointed to Probationary Constable status and subsequent permanent appointment – appear to have been subject to robust processes and are considered to be sound and reliable. While there are some gaps in the rationale supporting pass marks, links

with measures of performance and ongoing review, these tools are considered capable of supporting permanent appointment decisions.

### **Consistency of application**

The third question considered was whether “Comprehensive and sound procedures have been used to ensure consistent application of assessment methods”.

Five of the fourteen tools used to assess applicants (i.e. prior to someone entering the Police College) lack sound mechanisms to ensure nationally consistent administration, scoring and interpretation of results, reducing the ability of these tools to ensure that the quality of Recruits is nationally consistent. This may contribute to some variability in the quality of people entering the Police College from different districts.

In contrast, the assessment tools used post-recruitment (i.e. appointment to Probationary Constable status and subsequent permanent appointment) have sufficient mechanisms to ensure consistent application across the country.

We also note that assessment data recording errors and the lack of a central recording mechanism limit the ability of Police to accurately monitor assessment trends and to conduct empirical research to understand and improve the effectiveness of the assessment tools.

### **The reasoning ability of Recruits**

In an adjunct to this advice, Professor John Hattie was asked to examine whether or not there has been a reduction in the reasoning ability of recent Recruits. Professor Hattie states:

“From this review, it is concluded that while there is evidence that the minimum entry criteria on the Reasoning tests has declined, there is no evidence to believe that the overall performance of graduates from the training program has changed. It is not clear that any decline is attributable to the change in assessment, but is more likely related to decisions about the standards/ Guidelines/Rules of entry.” (2007, p. 26)

Professor Hattie’s full report is given in the Appendix.

### **Conclusion**

On balance, the assessment processes used by Police seem capable of ensuring that those achieving permanent appointment as Constables are able to perform the job. Primarily this is because of the multiple hurdle approach adopted by Police, whereby weaknesses in any one aspect of the assessment process are tempered by other, subsequent, parts of the process. Indeed, the demanding and broad range of tests that aspiring Constables must pass are significantly greater than those required for entry to most New Zealand organisations.

The weaknesses identified are more likely to have an impact on the overall efficiency, rather than effectiveness, of the appointment process. Potential impacts include devoting resources to people who will ultimately fail, devoting additional resources to people who take longer to train than would normally be expected, and also missing out on potentially suitable Recruits. Conducting a systematic and rigorous analysis of requirements, tightening some of the assessment methods and enhancing the consistency of application are likely to increase the focus of the recruitment process and, potentially, reduce costs.

Selecting staff is not an exact science. Police can improve, but they are not doing badly.

## 2. Contents

<b>1. Executive Summary .....</b>	<b>1</b>
Conclusion .....	2
<b>2. Contents .....</b>	<b>3</b>
<b>3. Approach .....</b>	<b>5</b>
<b>4. Baseline Practice Standards.....</b>	<b>7</b>
1. Determining minimum standards .....	7
2. Determining assessment methods .....	9
3. Consistent application .....	12
<b>5. Selection Process .....</b>	<b>13</b>
Recruitment.....	15
Initial Application .....	15
Testing Day.....	15
Background Enquiries .....	15
Selection Decision.....	15
Recruit Training.....	16
Summative Assessments .....	16
Probationary Constable Workplace Assessment Programme (PCWAP).....	16
Standards Assessment .....	16
Other PCWAP Assessments .....	17
<b>6. Identifying Role and Person Standards.....</b>	<b>18</b>
Role requirements .....	18
Knowledge, skills, attributes and other personal characteristics (KSAOs) required .....	19
Job analysis processes .....	20
Summary .....	20
<b>7. Assessment Methods .....</b>	<b>22</b>
Assessment coverage .....	22
Recruitment assessment.....	23
Health Questionnaire .....	23
Offences Clearance Check.....	24
Swimming Certificate of Competency .....	24
Cognitive Ability Testing (GRT2).....	25
Personality Testing (15FQ+) .....	27
Physical Appraisal Test (PAT).....	28
Behavioural Interview .....	29
Physical Competence Test (PCT).....	30
First Aid Certificate .....	31
Defensive Driving Certificate .....	32
Typing and Computer Skills Test .....	32
Reference Checks and Home Visit.....	33

SCOPE.....	34
Final Medical Clearance.....	34
Recruit training assessment.....	35
Written Summative Assessments.....	35
Practical Summative Assessments.....	37
Driver Training.....	39
Firearms Training.....	40
Defensive Tactics.....	41
Workplace Assessment Programme.....	42
LEGL 114: Introduction to criminal law and problem solving.....	44
Summary evaluation of assessment methods.....	46
<b>8. Consistency of Practice.....</b>	<b>48</b>
Recruitment assessment.....	48
Health Questionnaire.....	48
Offences Clearance Check.....	48
Swimming Certificate of Competency.....	49
Cognitive Ability Testing (GRT2).....	49
Personality Testing (15FQ+).....	50
Behavioural Interview.....	51
Physical Competence Test (PCT).....	52
First Aid Certificate.....	52
Defensive Driving Certificate.....	53
Typing and Computer Skills Test.....	53
SCOPE.....	53
Reference Checks and Home Visit.....	54
Final Medical Clearance.....	54
Recruit training assessment.....	55
Written Summative Assessments.....	55
Practical Summative Assessments.....	56
Driver Training.....	56
Firearms Training.....	57
Defensive Tactics.....	57
Workplace programme assessment.....	58
Standards Assessment.....	58
LEGL 114: Introduction to criminal law and problem solving.....	59
Summary evaluation of consistency of application.....	60
<b>9. References.....</b>	<b>62</b>
<b>10. Appendix.....</b>	<b>67</b>

## 3. Approach

The Honourable Annette King, Minister of Police, requested Dr Dave George of Cerno Limited to provide independent expert advice on the standards and assessment practices of the New Zealand Police, as used to assess applicants and sworn staff up to Probationary Constable level.

The scope of the Advice covers:

- › The processes Police use to determine minimum standards (e.g. knowledge, skills, abilities and other characteristics) for Police applicants, Recruits and Probationary Constables.
- › The processes Police use to determine the assessment methods (i.e. the tools and techniques) used to assess Police applicants, Recruits and Probationary Constables.
- › The procedures Police use to determine that the application of assessment methods is consistently applied nationally.

The approach used to provide this Advice involved four main stages:

- › **Stage One: Determine baseline practice standards**

Baseline standards were developed which represented standards of practice that could be reasonably expected of a New Zealand organisation when deciding on the assessment of applicants and trainees for appointment. Research was conducted to identify practice standards and guidelines endorsed by professional bodies (e.g. the New Zealand Psychological Society, the International Testing Commission and the American Psychological Association), the public sector in New Zealand (e.g. State Services Commission) and public sectors internationally (e.g. United States Department of Labor). These baseline practice standards were then used as a benchmark against which to compare Police practice.

- › **Stage Two: Document Police assessment standards and methods**

Based on the baseline practice standards determined in Stage One, requests were made to Police for documentation describing the processes used, people involved and outputs produced when determining minimum standards, and of assessment methods for evaluation of Police applicants, Recruits and Probationary Constables. Documentation was also sought to enable an evaluation of the consistency with which assessment methods were applied nationally. Where required, supplementary questions were identified, based on the information provided and discussions conducted with Police and assessment providers to clarify information received.

› **Stage Three: Compare Police assessment standards and methods with baseline practice standards**

The information received was used to compare Police assessment practices with the identified baseline practice standards and to describe those aspects that were consistent with the baseline standards and those that were not.

› **Stage Four: Report advice**

The information gathered in Stages One, Two and Three was compiled and analysed in this report, which describes the baseline practice standards, the level of consistency between Police practices and baseline practice standards, and the areas of Police practice requiring remedial attention.

› **Investigation into the reasoning ability of Recruits**

In addition to this comparison, an investigation was conducted by Professor John Hattie to establish whether or not there has been a reduction in the quality of recent Recruits during training. Professor Hattie examined the views and evidence provided by Senior Sergeant Iain Saunders RNZPC and Wayne Annan, General Manager Human Resources and provided a paper outlining his conclusions and recommendations, which is included in the Appendix.



## 4. Baseline Practice Standards

The baseline standards described in this section are derived from practice standards and guidelines endorsed by a range of professional bodies. A high level of consistency was observed in the practice standards reviewed. The guidelines referred to were endorsed by the New Zealand Psychological Society, the Australian Psychological Society and the American Psychological Association (American Educational Research Association AERA, American Psychological Association, & National Council on Measurement in Education, 1999; Australian Psychological Society, 2002; Kendall, Jenkinson, de Lemos & Clancy, 1997; Love & Whittaker, 1997). The baseline standards listed here represent the central themes found in all three sets of guidelines. They are consistent with, although more comprehensive than, the guidelines produced by the International Testing Commission (which focuses on the use of formal psychometric tests; International Testing Commission, 2000). Guidelines endorsed by professional bodies in other countries (such as Canada and Britain), or that are included in the professional literature, were also researched. Standards or findings that are specific to these additional sources are referenced in the text.

### 1. Determining minimum standards

All guidelines reviewed stated the indispensability of a systematic and robust analysis of job requirements, and of the personal characteristics required to perform the role to an acceptable standard, as the basis of any assessment in a work setting. Human resource practitioners and organisational psychologists regard a sound definition of role requirements as the foundation of any selection process, which “ultimately determines the effectiveness of a staffing programme” (Ployhart, Schneider, & Schmitt, 2006, p. 222). Without specifying role requirements and KSAOs, it is not possible to make an informed decision about the assessment procedures that are best suited to selecting high performing staff members.

As well as being an accepted baseline standard for human resource practice, defining sound role requirements is a specific expectation of New Zealand State Sector organisations. Both the State Sector Act 1988 and the Police Act 1956 require that Public Service employers appoint employees on merit, where merit is defined as “the person who is best suited to the position”. The State Services Commission (SSC) interprets this as requiring “that merit (or the criteria on which an appointment will be made) is carefully defined in job specifications. The definition of merit is not fixed but is related to the particular requirements of a specific position. It usually refers to the qualifications, experience and personal qualities, all of which must be relevant to the position” (SSC, 1998, p. 13). Meeting this requirement is dependent on a sound definition of the role requirements.

Therefore, the first baseline standard and accompanying criteria require that:

**A systematic and robust process was used to identify the minimally acceptable knowledge, skills, attributes and other personal characteristics (KSAOs) required of Police Recruits, Probationary Constables and Constables. This process:**

- 1.1. Documented the key objectives, duties and work activities (i.e. role requirements) that Probationary Constables and Constables are expected to perform competently (i.e. without support).
- 1.2. Documented the different contexts in which Probationary Constables and Constables are expected to perform competently. For example, level of contact with the public, nature and frequency of conflict situations, and level of support.
- 1.3. Documented the range of policing environments in which Probationary Constables and Constables are expected to perform. For example, road and community policing environments, and metropolitan and rural environments.
- 1.4. Documented any regional differences in activities and context.
- 1.5. Documented any differences between Probationary Constables and Constables in terms of their role and performance requirements.
- 1.6. Provided reasonable documented justification that individuals possessing the identified personal characteristics perform more effectively on the job than individuals lacking such personal characteristics.
- 1.7. Documented any differences between Police Recruits, Probationary Constables and Constables in terms of the required knowledge, skills, attributes and other personal characteristics.
- 1.8. Documented which knowledge, skills, attributes and other personal characteristics Police Recruits, Probationary Constables and Constables would be expected to develop through training and on-the-job experience.
- 1.9. Involved a range of people who are suitably knowledgeable about the roles (e.g. Probationary Constables and Constables) to which the results have been generalised.
- 1.10. Used role requirements to validate the minimally acceptable level of each area of knowledge, skill, attribute and other personal characteristic required.
- 1.11. Was regularly updated to take into account any changes in organisational needs, technology, equipment, work assignments and work environments.

The above criteria as listed can be apportioned into three main groups:

- > Criteria numbered one to five describe the nature and breadth of the role requirements including the key objectives, duties and work activities of the role;
- > Criteria six to eight highlight the person requirements such as type and quantity of knowledge, skills, attributes and other personal characteristics (KSAOs) required of the incumbent; and
- > Criteria nine to eleven emphasise the importance of following a rigorous job analysis process, to ensure the role requirements and KSAOs thus documented are accurate.

The initial group of five criteria focus on accurate documentation of the range of role requirements. The breadth of contexts (i.e. Criterion 2) provides information on the complexity of demands involved. The different environments or Regional settings within which incumbents work (i.e. Criteria 3 and 4) reflect the range of tasks that may be involved. Criterion 5 takes into account the differences in tasks and demands at different stages of the selection process, information that helps verify what will be expected of incumbents at each stage of the process.

The next group of criteria consider the KSAOs required of the person. Criterion 6 ensures that the KSAOs documented are related to performance on the job and Criterion 7 emphasises the need for any differences in role requirements to be reflected as differences in KSAOs. Criterion 8 ensures clarity as to which KSAOs applicants need to possess when they are recruited, and which KSAOs are trained for at different stages of the recruitment process.

The final group of criteria highlight the importance of the rigour of the job analysis process (Schippmann et al, 2000; Society of Industrial and Organisational Psychology [SIOP], 2003). Criterion 9 ensures that the role requirements and KSAOs are balanced and complete, by including a range of informed viewpoints in their construction. Criterion 10 ensures that the minimum standards set for KSAOs are independently verified and the last criterion highlights the importance of job and person requirements being kept up-to-date with the realities of the role.

## 2. Determining assessment methods

All of the guidelines reviewed specified the need for the assessment methods to be chosen using a systematic and robust process. This baseline standard ensures that Recruits are selected on the basis of accurate measures of their suitability for the job, to prevent the selection process being unduly influenced by factors that are irrelevant to their capacity to do the job well. The need for this baseline standard is underscored by the fact that it is common for organisations to select staff using methods that are *not* effective predictors of job performance (Robertson & Smith, 2001).

Therefore, the second baseline standard and accompanying criteria require that:

**A systematic and robust process was used to identify the assessment methods.  
This process needs to provide reasonable documented justification that:**

- 2.1. The range of assessment methods used covers all the minimally acceptable knowledge, skills, attributes and other personal characteristics required of Police Recruits, Probationary Constables and Constables.
- 2.2. Each assessment method is reliable and valid for the intended purpose.
- 2.3. Assessment cut-scores and/or pass marks are soundly based and reflect the minimally acceptable knowledge, skills, attributes and other personal characteristics required of Police Recruits, Probationary Constables and Constables.

2.4. Assessment administration policies and guidelines enable accurate replication of the required administration conditions, and the scoring, interpretation and communication of assessment results.

2.5. All reasonable efforts have been made to reduce the potential discriminatory impact of assessment tools and procedures.

2.6. Assessment review policies and processes are sound and responsive to changes in role expectations or when assessment methods are compromised.

Criterion 1 emphasises the importance of the assessment process covering the range of KSAOs required, prioritised by their impact on job performance, because if a particular critical KSAO is unmeasured, then applicants may be selected who cannot perform the role to an acceptable standard. This linkage of assessment methods to the range of KSAOs is essential to choosing Recruits that will perform well on the job (AERA, 1999; SIOP, 2003), and also allows Police to fulfil their legal obligation to select staff on merit.

Criterion 2 ensures that the assessment methods are reliable (i.e. consistent) and valid in that they measure what they are intended to measure.

- › Reliability is measured by examining whether an assessment method gives consistent results. For example, two interviewers make separate assessments of a group of applicants. If one interviewer identifies a certain individual as the most suitable applicant, and another identifies her as the least suitable applicant, then they cannot both be correct. This would be an indication of low reliability. The same would apply to a test of reading skills that has two different versions. If applicants usually obtain similar scores on the two versions of the test, this would be evidence of reliability. If applicants often obtain very different scores on the two versions of the test, this would be evidence of unreliability.
- › An assessment method is valid if it measures what it is intended to measure. An assessment method can be reliable but not valid. For example, an organisation might wish to assess physical fitness. If the only assessment method used was body mass index (weight divided by height squared), this would produce reliable results (because weight and height can be accurately measured), but the results would probably be of low validity (because body mass index has only a limited relationship with overall physical fitness).

There are a number of possible ways of demonstrating the validity of an assessment method, and no one way is prescribed (AERA, 1999). In general, validity is measured by examining whether the results of an assessment either accurately predict future performance (predictive validity), or can be seen as a sample of job performance (content validity). Validity can be established either through internal research in the organisation, or reference to external research in other settings, but in either case the evidence must be sound (AERA, 1999). Evaluating the soundness of validity evidence requires professional judgement. For example, a predictive validity study compares scores on the selection test to measures of job performance. To give useful results in the Police context, it must:

- › Involve sufficient numbers of Recruits to be statistically reliable.

- › Be well designed, so that spurious factors do not affect the results (e.g. if job performance is rated by the Recruits' managers, and managers know Recruits' selection test scores, then managers' ratings may be influenced by this knowledge. This would create a false impression that the test scores are predicting performance).

Similarly, validity evidence based on test content must be thorough and explicit (AERA, 1999). The validity of credentialing tests (e.g. qualifying tests such as those a Probationary Constable must pass to become a Constable) is usually examined by examining their content validity. As will be described in the body of the report, a number of the Police assessment processes for Recruits and Probationary Constables fall into this category. Valid credentialing tests typically accurately reflect the content of the role or task that they are assessing for and are pitched at a level of difficulty that reflects the skill level required for satisfactory job performance (AERA, 1999).

Criterion 3 underscores the impact the choice of cut-off scores has on selection outcomes, and thus that this choice should not be made arbitrarily (AERA, 1999). If a cut-off score is set too low, Recruits will be selected that are not able to perform the role to an acceptable standard. If a cut-off score is set too high, some applicants who could have performed the role to an acceptable standard will not be selected. Where feasible, cut-off scores should be supported by empirical data that shows the relationship between assessment scores and performance (AERA, 1999). In some circumstances, it may not be feasible to gather empirical data to support a choice of cut-off score. In these cases, it is still essential to set cut-off scores using a well-reasoned, documented and replicable process (AERA, 1999). This will often be an interim measure while gathering empirical data to support a choice of cut-off score. Cut-off scores should be set to ensure that Recruits are capable of satisfactory job performance (rather than having a pass mark that is adjusted to regulate the number or proportion of people passing the test).

Criterion 4 is a prerequisite for ensuring that the assessment procedure is controlled and standardised, to prevent the occurrence of unnecessary measurement errors. The need for this criterion is underscored by the frequent occurrence of errors in test administration, scoring and interpretation in real-world assessment situations (Kendall et al, 1997). For example, test administrators must not give some applicants an advantage by allowing extra time to complete assessments, or by scoring answers as correct where the scoring guidelines classify them as incorrect.

The importance of criterion 5 is emphasised by the provisions of the Human Rights Act 1993 that prohibits unlawful employment discrimination on each of 13 grounds (SSC, 1998). All of the guidelines reviewed also specified an ethical requirement for organisations to reduce the potential discriminatory impact of assessment procedures, even against groups that are not named in the Act.

Criterion 6 ensures that the assessment process remains effective over time. For example, if the use of computers becomes an essential part of a role, then it may become worthwhile for an organisation to start assessing computer skills as part of the selection process.

### 3. Consistent application

The final baseline standard ensures that the assessment is conducted in a standardised and accurate manner. All of the guidelines reviewed specified that consistent application of assessment methods is essential.

Therefore, the third baseline standard and accompanying criteria require that:

**Comprehensive and sound procedures have been used to ensure consistent application of assessment methods. Procedures need to ensure that:**

- 3.1. Individuals are appropriately qualified and competent in the use of the assessment tools for which they are responsible.
- 3.2. The integrity of assessment materials and confidentiality of assessment results are protected.
- 3.3. Reassessment policies and procedures are based on a sound rationale.
- 3.4. Mechanisms are in place to ensure that administration, scoring, interpretation and communication of results is nationally consistent and matches recognised standards of assessment practice.

Criterion 1 ensures that the individuals who administer and interpret assessments are able to do so in an appropriate manner. It is common for untrained individuals to make a number of errors in the administration, scoring and interpretation of tests and assessments (Kendall et al, 1997). These errors can result in the selection of applicants who do not have the ability to perform the role to an acceptable standard, and in other unintended consequences.

Criterion 2 is critical in ensuring the continued validity of the assessment process. If a test is made public, and some applicants are able to practice tests and assessments before the assessment day, they may be unfairly advantaged. This criterion also ensures the privacy of applicants.

Criterion 3 considers those occasions where applicants obtain assessment results that indicate they are currently unable to meet minimum performance standards. Organisations may choose to allow reassessment, i.e. the applicant is allowed to re-sit the test at a later date. This procedure is justified in some circumstances (e.g. an applicant is allowed to re-sit a physical fitness test after completing physical fitness training, or to complete a qualifying exam after further study), but in other circumstances it may create perceptions of unfairness or lax standards. This criterion ensures that policies for allowing applicants to re-sit tests are sound, clear and consistently applied.

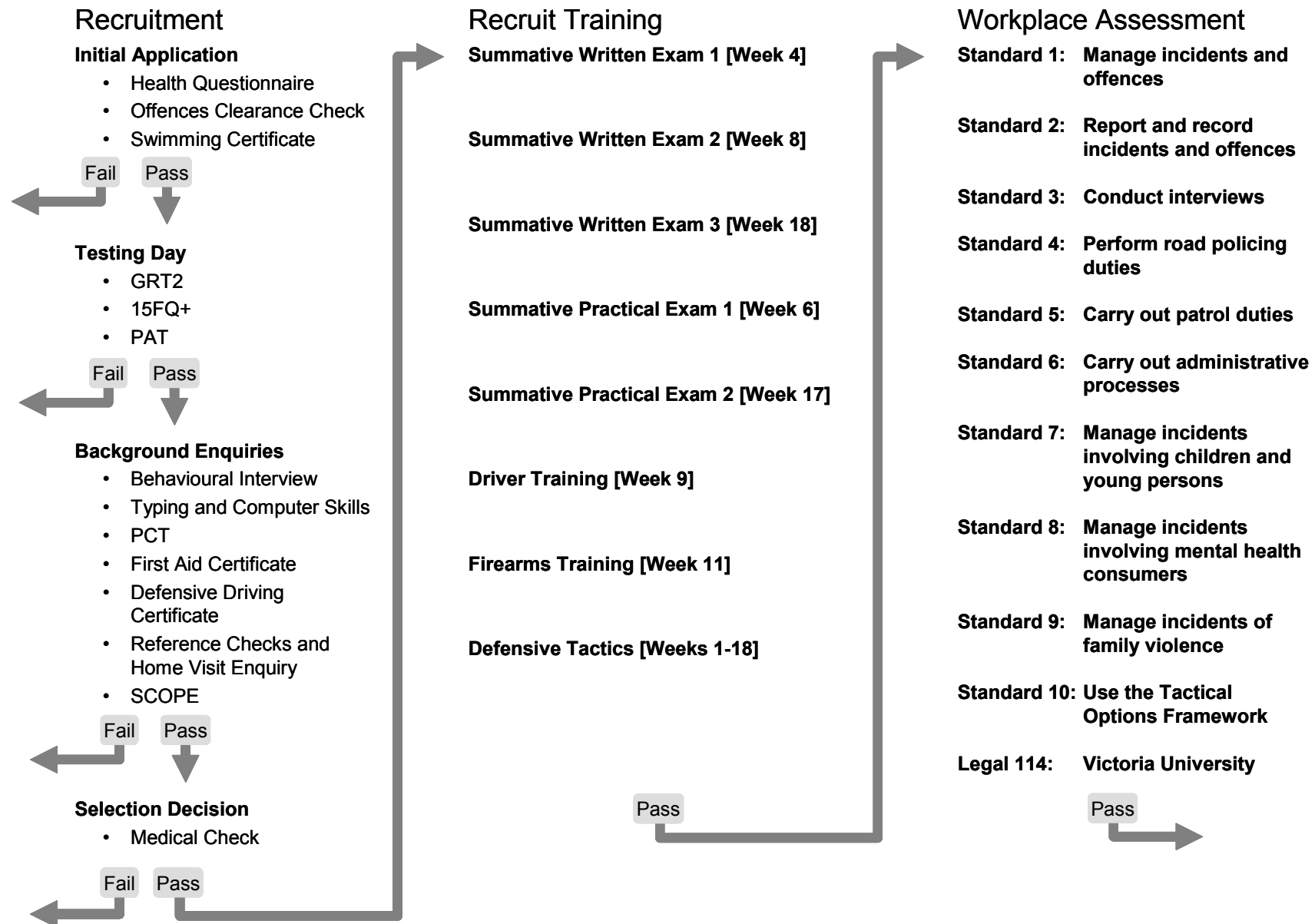
Criterion 4 ensures that the assessment procedure is consistently implemented as intended, to prevent the occurrence of unnecessary measurement errors. The importance of this criterion is underscored by the frequent occurrence of errors in test administration, scoring and interpretation in real-world assessment situations (Kendall et al, 1997).

## 5. Selection Process

The selection process used by the New Zealand Police to identify, assess and train Constables takes an individual between two and three years to complete. During this time, an individual can progress from being:

- › An applicant during Recruitment;
- › A recruit completing Recruit Training at the Royal New Zealand Police College (RNZPC); and
- › A Probationary Constable completing the Workplace Assessment Programme.

The time taken to complete the selection process through to becoming a permanently appointed Constable is influenced by a range of factors, e.g. how long it takes for an individual to meet the standards of the specific knowledge, skill, ability and other characteristics (KSAOs) required (e.g. fitness standards), and the availability of a place on the RNZPC Recruit Course. During this process, individuals are regularly assessed and need to meet specific standards to progress. The assessments used in each of the three phases are shown in the diagram on the following page.





## Recruitment

The recruitment process involves four main stages (i.e. Initial Application, Testing Day, Background Enquiries and Selection Decisions) requiring an applicant to 'pass' each stage before they can continue to the next. Following successful completion of the final fourth stage, the District Recruitment Officer recommends that an offer of employment be made. The final recommendation is based on a review of the applicant's performance throughout the recruitment process.

The assessment tools used during the recruitment process are listed as follows:

### Initial Application

Applicants typically attend a recruitment seminar and hear a presentation about what is involved in a career with the Police, and the procedures involved in the recruitment process. The key assessment tools used in the Initial Application are:

- > A Health Questionnaire (as the first stage of the applicant's medical assessment);
- > An Offences Clearance Check; and
- > A Swimming Certificate of Competency that applicants are required to provide.

### Testing Day

Invited applicants attend a Testing Day where they complete:

- > Cognitive ability tests (GRT2);
- > A personality test (15FQ+); and
- > A Physical Appraisal Test (PAT).

### Background Enquiries

During the Background Enquiries stage, applicants:

- > Attend a Behavioural Interview;
- > Complete a Physical Competency Test (PCT)
- > Provide a First Aid Certificate;
- > Provide a Defensive Driving Certificate;
- > Complete a Typing and Computer Skills Test; and
- > Complete a work experience exercise, titled SCOPE.

It is during this stage that the District Recruiting Officer conducts referee checks and a home visit.

### Selection Decision

During the Selection Decision stage, the District Recruiting Officer considers all the assessment information gathered on an applicant during the recruitment process. Where the District Recruiting Officer considers an applicant to be a clear 'yes' or 'no', they make an employment recommendation to the District Human Resources Manager. Where the District

Recruiting Officer considers an applicant NOT to be a clear 'yes' or 'no', the decision is referred to the District HR Manager. If an applicant is considered suitable, they are placed on the National Waiting List for inclusion in a Recruit Training Course, contingent on them successfully obtaining a Final Medical Clearance and having a 'current' PCT (i.e. successfully completed within six months of the beginning of their Recruit Course).

## Recruit Training

On acceptance for Recruit Training, a successful applicant receives a 'call up letter' to schedule their attendance at the Recruit Training Course. The successful applicant needs to have successfully completed a PCT no longer than six months in advance of the commencement of their training. On arrival at the Royal New Zealand Police College (RNZPC), the new recruit signs their call up letter as formal acceptance of employment and is sworn in as a member of the New Zealand Police.

### Summative Assessments

Recruit performance on the 19 week Recruit Training Course is assessed in eight separate assessments. Three of the assessments are Written Summative Examinations, covering knowledge of a range of policing topics (e.g. the law, policies and procedures for dealing with suspects, and Maori responsiveness). Two of the assessments are Practical Summative Assessments, in which Recruits must handle simulated scenarios.

The remaining three assessments examine operational skills involving Driver Training, Firearms Training and Defensive Tactics.

On successful completion of the Recruit Training Course, the Recruits graduate, are appointed as Probationary Constables and posted to a District to commence general duties.

## Probationary Constable Workplace Assessment Programme (PCWAP)

The PCWAP consists of ten PCWAP Standards and a university paper (LEGL 114). Probationary Constables are required to complete the ten PCWAP Standards within two years of graduation. On a day-to-day basis, the PCWAP is managed by District Workplace Assessors (DWA) and supported by Field Training Officers (FTO) who assist and assess a Probationary Constable using a structured assessment process including defined performance criteria (Royal New Zealand Police College, 2007).

### Standards Assessment

The PCWAP Standards assessments require Probationary Constables to demonstrate abilities in a range of areas. The Probationary Constable compiles evidence and demonstrates their competence on each Standard for assessment by a Designated Assessor. When a Standard is not met, a Probationary Constable is permitted two further reassessments; if these attempts are also unsuccessful, a performance management process is implemented. The Standards are:

- › Standard 1: Manage incidents and offences
- › Standard 2: Report and record incidents and offences
- › Standard 3: Conduct interviews
- › Standard 4: Perform road policing duties
- › Standard 5: Carry out patrol duties
- › Standard 6: Carry out administrative processes
- › Standard 7: Manage incidents involving children and young persons
- › Standard 8: Manage incidents involving mental health consumers
- › Standard 9: Manage incidents of family violence
- › Standard 10: Use the Tactical Options Framework

### **Other PCWAP Assessments**

Probationary Constables are also required to complete a university paper (LEGL 114: Introduction to criminal law and problem solving) and achieve a satisfactory rating in their current performance appraisal.

Upon the successful completion of the ten PCWAP Standards, LEGL 114 and a satisfactory performance appraisal, a Probationary Constable attains 'Permanent Appointment' as a Constable.

## 6. Identifying Role and Person Standards

This section describes and reviews the evidence for consistency with the first Baseline Standard.

**A systematic and robust process was used to identify the minimally acceptable knowledge, skills, attributes and other personal characteristics required of Police Recruits, Probationary Constables and Constables.**

The criteria specifying what is considered a systematic and robust process is detailed in the section titled, 4. Baseline Practice Standards and involves three main themes.

The process of identifying role and person standards needs to:

- > Accurately describe the nature and breadth of the role requirements including the key objectives, duties and work activities of the role;
- > Accurately identify the person requirements such as type and quantity of KSAOs required of the incumbent; and
- > Use a rigorous job analysis process.

Following are the findings derived from documentation provided by Police, and discussions with Police National Headquarters Human Resources staff, District Recruitment Officers, Physical Education Officers, and Royal New Zealand Police College managers and instructors.

### Role requirements

#### Meets standards

The relevance of training to operational requirements is reviewed twice a year through the Curriculum Reference Group, which is tasked with seeking National and District perspectives on training to “ensure courses meet the operational and policy needs of the organisation, the professional development needs of individual Police Officers and demonstrate best practice in all areas” (Initial Training Group, 2005a, p. 1). While this process provides a range of information to assist training staff to adapt programme content, it is not clear how this information is used to review or modify assessment processes.

#### Scope for improvement

While a number of recent, detailed job descriptions describing the purpose and key accountabilities for common Constable roles are in everyday use (e.g. Constable: General Duties, Traffic Constable and Traffic Alcohol Group Constable), there is no evidence to show that a systematic and robust process was used to develop these descriptions.

The job descriptions Police have developed to describe the different demands of the range of policing environments for Constable positions appear limited to those noted above. While as would be expected, the objectives and duties of each were clearly different, no evidence was available to describe the rationale used to define those differences. In many cases the

descriptors used were not sufficiently specific to enable comparisons to be made, e.g. “Maintaining personal, physical and emotional health” (New Zealand Police, 2001a, p. 2) and there was no indication of the frequency and importance of different job requirements.

It also appears that Police have little documented information on the range of contextual factors that impact on role requirements (e.g. role contexts, role complexity and regional differences) or on the differences in expectations between Probationary Constables and permanently appointed Constables (P. Harris, personal communication, 17 September, 2007).

Police have conducted job evaluation exercises that have involved managers in developing the description of key tasks and levels of responsibility held by Constables (New Zealand Police, 1998). However, these processes were focused on the factors that underpinned the relative worth of jobs, and did not provide a comprehensive description of the range of key objectives, duties and work activities.

## **Knowledge, skills, attributes and other personal characteristics (KSAOs) required**

### **Meets standards**

Studies conducted by Police have determined that individuals possessing some specific KSAOs are likely to perform more effectively on the job than individuals lacking such characteristics.

- > Cognitive ability has been shown to be a useful predictor of Recruit and Police Officer performance in New Zealand (Black, 1995; Duggan, 2000; O’Callaghan, 2003).
- > Significant relationships have been found between certain personality traits and recruit performance (Black, 1995).
- > A review of physical fitness assessments was conducted by specialists using the input of a range of experts (i.e. Police Medical Advisor, physical education specialists and instructors), international research, fitness requirements applied in other Police jurisdictions and Otago University’s Human Performance Centre’s Database to determine physical fitness requirements and levels (Wilson & Toomey, 1993, Wilson & Toomey, 1990).
- > A medical specialist was contracted to review the visual acuity, hearing, epilepsy, diabetes, asthma, and hypertension standards expected of Recruits, to enable them to meet the key functional demands of a Constable’s role (Robinson, 2006). The review was conducted by a highly experienced medical specialist, who used a range of information from credible sources (e.g. specialist ophthalmic advice, medical research, and health factors considered by comparable agencies such as the New Zealand Defence Force, the New Zealand Fire Service and Police Services in the United Kingdom) and ensured recommendations complied with statutory requirements (e.g. Land Transport Act 1998).
- > Specific examples of situations where specific KSAOs have a direct relationship with operational requirements of Constables include Defensive Tactics (Defensive Tactics Section, 2003) and Police Driving (Professional Police Driver Programme, 2005). While there is little evidence of how each of the KSAO descriptions were developed, the references to operational procedures in New Zealand and other jurisdictions (e.g.

Training Service Centre, 2001) and to legal requirements (e.g. Evidence Act 2006 and Land Transport Act 1998), and the involvement of experienced Police Officers in compiling the training material, suggests that the descriptions are a reasonably accurate representation of the KSAOs trainees would expect to develop.

Clear and comprehensive descriptions of the key knowledge and skills Probationary Constables would be expected to develop through training and on-the-job experience are outlined in the current versions of Police training manuals and other programme materials (e.g. Initial Training Group, 2007a; Initial Training Group, 2007b; Recruit Training Group, 2006; Victoria Police Education Programme, 2007).

### **Scope for improvement**

The absence of a sound description of the key job requirements means that there is not a sound foundation on which to develop a defensible description of the range and relative importance of KSAOs for the different roles.

## **Job analysis processes**

### **Meets standards**

Police have appropriately responded to changes in legislation such as the Human Rights Act 1993 (e.g. Wilson & Toomey, 1993) and operational changes in policing practices such as vehicle pursuit (i.e. Carson, 2003). However, there is an absence of established policies or mechanisms to govern the regularity with which, or the instances when, a review of job requirements and KSAOs is necessary.

Police have responded to identified operational needs where specific KSAOs have been considered necessary for Police Officers to meet operational needs, for example:

- > Defensive Driving
- > Swimming Competency
- > First Aid
- > Typing and Computer Skills

### **Scope for improvement**

Established processes to regularly update job requirements and KSAOs to take into account any changes in organisational needs, technology, equipment, work assignments and work environments do not appear to be in place. The need for established updating processes has been recognised by advisors on medical standards (Robinson, 2006) and psychological testing (e.g. Barrett 2005b; O'Callaghan, 2003), and by reviewers of the relationship between recruitment and recruit training effectiveness (Trappitt, 2007) and the length of the Recruit Course (Anderson & Penny, 2003).

## **Summary**

Examination of practices show that Police have an accurate understanding of some specific job requirements (e.g. conduct criminal investigations, respond to traffic emergencies and

conduct lawful interviews) and a number of KSAOs (e.g. medical and physical fitness standards), and the processes used to identify the relevant information were sound.

However, no evidence was found of a systematic and robust process being used to determine the breadth of job requirements (i.e. objectives, duties and tasks) and the range and relative importance of KSAOs required of Recruits, Probationary Constables and Constables. In absence of such an analysis, it is not possible to state definitively that the minimally acceptable KSAOs Police are using for assessment accurately reflect the requirements of the roles. As a result, the minimally acceptable KSAO levels currently assessed may be greater or less than required. In addition, some key KSAOs currently being assessed may not be critical for successful performance and conversely, there may be KSAOs that are critical for successful performance that are not being considered when their inclusion would increase the accuracy of the selection process.

## 7. Assessment Methods

This section describes and reviews the evidence related to the second Baseline Standard.

### **A systematic and robust process was used to identify the assessment methods.**

The criteria specifying what is considered a systematic and robust assessment process in detailed in the section titled '4. Baseline Practice Standards' and involves six main themes.

The process of identifying the assessment methods used needs to:

- › Cover all the minimally acceptable knowledge, skills, attributes and other personal characteristics required;

And, ensure that:

- › The tools identified are reliable and valid for the intended purpose;
- › Cut-scores or pass marks are soundly based;
- › Administration policies and guidelines are clear and complete;
- › Reasonable steps have been taken to reduce potential discriminatory impact; and
- › Review policies and processes are sound.

Following are the findings derived from documentation provided by Police, and discussions with Police National Headquarters Human Resources staff, District Recruitment Officers, Physical Education Officers, and Royal New Zealand Police College managers and instructors.

### Assessment coverage

#### **Scope for improvement**

As Police have not conducted a systematic and robust analysis to identify the range of KSAOs required of Police Recruits, Probationary Constables and Constables, it is difficult to determine whether the range of assessment methods adequately covers the range of KSAOs required in these roles. Despite a number of the assessment tools having an obvious and direct relationship with important work behaviours and knowledge (e.g. the Probationary Constable Workplace Assessment Standards tests), attributes that are less apparent (e.g. tolerance for stress) require a systematic and robust process to establish defensible requirements. For Police to ensure all KSAO requirements are covered by the range of assessment processes, a systematic and robust analysis of those requirements is an essential first step.

Were defensible requirements established, suitable assessment methods could be 'mapped' against these requirements. The documentation provided shows that Police have made an attempt to summarise which KSAOs are assessed by their current range of assessment methods. However, it appears the summary is limited to applicant assessment (i.e. does not include Recruits or Probationary Constables) and is based on the judgements of a single Recruiting Officer (G. Van Ooyen, personal communication, 17 September, 2007). Thus, Police are unable to demonstrate that the range of assessment methods used covers all the



minimally acceptable knowledge, skills, attributes and other personal characteristics required of Police Recruits, Probationary Constables and Constables.

## Recruitment assessment

The recruitment process assesses applicants during the four stages of recruitment (i.e. Initial Application, Testing Day, Background Enquiries and Selection Decision). Below, each assessment method used is described along with the KSAOs it is intended to assess, how assessment results are used and how the assessment method compares with the baseline standards. Each assessment method is dealt with in the order typically encountered by those being assessed.

### Health Questionnaire

The Health Questionnaire and Final Medical Clearance are two stages in the assessment of an applicant's medical fitness (New Zealand Police, June, 2007a). As part of their initial application, an applicant completes a standardised Health Questionnaire, responding to questions regarding general health, hearing, conditions relating to keyboard and computer mouse use, mental health, asthma and vision. An applicant's responses are reviewed by the Recruiting Officer to ensure they are medically suitable. If no conditions are identified that would prevent them from performing full Police duties during an expected career of up to 32 years, they will continue on to the next stage of the recruitment process. Where applicants are obviously not medically suitable, they are not accepted for training and are informed immediately. If their health status has the potential to prevent them from serving as a sworn Officer, they are asked to obtain a specialist report investigating the health issues concerned, at their own expense. A decision as to whether they progress to the next stage is then made on the basis of the applicant's specialist report, by the Manager Recruiting and Appointments.

### Meets standards

The ability of the Health Questionnaire to produce reliable information is assisted by clear instructions, descriptions of the health issues and conditions on which information is requested and an emphasis on the benefits (e.g. assuring applicant healthy and safety) and consequences (e.g. false or misleading information will lead to an application being declined) of completing the form accurately (New Zealand Police, 2006a).

The validity of the Health Questionnaire requests for information appears soundly based, in that they are consistent with the requirements of the Final Medical Clearance (Robinson, 2006).

### Scope for improvement

In the information provided, there was an absence of mechanisms to enable regular review of the reliability and accuracy of applicant responses to the Questionnaire (e.g. consistency between Questionnaire responses and subsequent Final Medical Clearance findings) to allow identification of improvements.

## Offences Clearance Check

Typically at the conclusion of a recruitment seminar, applicants register their interest in applying to become a Police Officer by completing a Recruitment Registration Form, which includes a declaration of any previous convictions for offending, and an authorisation for Police to make any enquiries deemed appropriate to determine the applicant's suitability. If applicants are deemed suitable, they continue in the recruitment process.

### Meets standards

The reliability of the Offences Clearance Check is supported by documented standards that are soundly based on the policy requirements described in the Commissioner's Policy on Proven Charges.

The validity of the information provided by applicants is assisted by:

- > Checks made against Police intelligence information (W. Kennedy, personal communication, 25 September, 2007).
- > Monitoring and identifying trends in offences data such as identifying differences between the number of Officers with previous offences and those without who are involved in complaints made to the Police Complaints Authority (Annan, 2007b).

Guidelines provided to Recruits were considered sound. Clear instructions are provided to Recruits in recruitment registration forms and the Police Application Form (New Zealand Police, 2005a), in addition to a clear outline of the consequences of not disclosing information, or of providing false or misleading information.

Clear procedural guidelines describing how an Offences Clearance Check should be conducted have been developed to assist staff responsible for conducting the assessment to follow sound and complete processes (New Zealand Police, 2007d).

## Swimming Certificate of Competency

The Swimming Certificate of Competence tests swimming ability. Applicants are required to swim 50 metres (in 54 seconds or less), tread water in the same spot (for at least 5 minutes), and 'duck dive' to retrieve and swim with an object in 3 metres of water while being observed by a certified Police Assessor. Testing for the Certificate is typically carried out by arrangement with a Police assessor prior to the Testing Day and an applicant produces their Certificate on their Testing Day. Applicants must meet competence requirements to proceed to the next stage of the recruitment process.

### Meets standards

The reliability of the swimming assessment is assisted by the use of specified assessment procedures and certification standards.

The validity of the requirement for a Swimming Certificate of Competence is assisted by an obvious relationship with operational needs (i.e. to ensure Police Officers have the ability to safely enter the water if required in the course of their duties).

The minimum standard that applicants must achieve is specific and clear for applicants and for assessors.

Administration instructions were clear and specific, supported by forms to collate and score applicant performance.

### **Scope for improvement**

The process used to establish and justify the standards required for a Swimming Certificate of Competence was not clear from the documentation provided.

Mechanisms to enable regular review of the reliability and validity of swimming assessments were not described in the information provided (e.g. the relationship between Certification and the swim survival practical assessment, which is part of Defensive Tactics Training conducted on the Recruit Course).

### **Cognitive Ability Testing (GRT2)**

In 1994, New Zealand Police initiated research to evaluate and expand their recruit psychological testing programme (e.g. Black, 1995; Duggan, 2000), culminating in an in-depth review of alternative instruments to evaluate the suitability and effectiveness of current tests compared to alternatives available, and to examine the potential advantages of Internet-based testing (O'Callaghan, 2003). As a result of that review, a paper-based version of the GRT2 cognitive test battery was introduced in 2004 (along with the 15FQ+ personality test described below).

The GRT2 sub-tests include:

- > Abstract Reasoning - tests the ability to learn innovative or new information. It requires the identification of patterns presented in diagrammatic form (so avoiding dependency on learned verbal or numerical skills).
- > Numerical Reasoning - assesses the ability to think and reason using numbers. It requires the solution of problems ranging in complexity from simple addition, subtraction, multiplication or division operations to problems involving number sequences and applied mathematical problems in work situations.
- > Verbal Reasoning - assesses the ability to think and reason using language. It requires the solution of problems involving language presented in a number of forms (e.g. tables, diagrams, flowcharts and plain text), including defining words and providing synonyms, identifying the underlying meanings of common words and phrases, and answering comprehension questions from a passage of prose.

The most recent description of how the test results are used requires applicants scoring Stanine 3 or less on any of the three sub-tests to be 'stood down' and asked to come back when better prepared. If an applicant scores Stanine 4 on any sub-test, 'other' non-test variables are considered before asking the applicant to proceed or to 'stand down' (New Zealand Police, 2006b).

### **Meets standards**

Reliability of the GRT2 was considered sound because:

- > A research study to test the instrument's psychometric properties described the GRT2 as reliable and measuring separate, though correlated, cognitive ability dimensions. The researcher concluded, "The GRT2 was judged to have adequate psychometric properties for applicant screening and selection purposes" (Chernyshenko, 2005, p3).

The validity of the GRT2 for the assessment of cognitive ability is supported by:

- > Evidence for construct validity and predictive validity for samples of different occupational groups from the United Kingdom, South Africa and New Zealand (Psytech International Limited, 2007).
- > A validity study involving New Zealand Police Recruits demonstrated a significant relationship between GRT2 scores and scores on the Final Summative Written Assessment (which is completed at the conclusion of training; Barrett, 2005b). Earlier research has also supported a relationship between cognitive test results and training performance (Black, 1995; Duggan, 2000). However, as mentioned below, a sound criterion accurately describing on-the-job performance is needed before evidence for the GRT2's predictive validity can be extended beyond training success to include job performance.

Assessment administration policies and guidelines are considered sound because:

- > Test administration instructions are clear and specific.
- > Supporting technical documentation is comprehensive and clearly described, including descriptions of instrument construction and specifications, interpretation strategies, aptitude domains measured, and psychometric properties including reliability, validity and normative data (Psytech International Limited, 2007). The GRT2 has been standardised on a sample of 5,183 New Zealanders of working age. Additional information is also available on the publisher's web site (<http://www.psytech.co.uk/tests/grt2.htm>).
- > Implementation of the GRT2 was systematically planned and methodical, and included strategies for a phased 'change-over' of tests, customising of reporting procedures, administration and interpretation training (OPRA Consulting, 2004a).

The Police have taken reasonable steps to reduce potential discriminatory impact. Following a study to examine such bias, Chernyshenko (2003) described the GRT2's sub-tests as displaying no evidence of measurement bias across gender and ethnic groups (i.e. European and Maori/Pacific Island People).

### Scope for improvement

Examination of the empirical validity of the GRT2 as a predictor of Police Officer performance is limited by the lack of a sound criterion reflecting 'actual on-the-job performance'. Planning for empirical research that establishes such a criterion has been proposed in the past (e.g. Barrett, 2005a, 2005b; OPRA Consulting, 2004a), but no evidence was observed of progress on this issue.

The justification for the GRT2 cut-score was not sufficient to meet baseline standards. Reviewers have criticised the cut-off scores used by Police for a lack of basis in empirical evidence relating test scores to actual performance (Barrett, 2006), and making arbitrary changes to cut-scores rather than using a standard setting process that is systematic, credible and replicable (Hattie, 2007). Hattie strongly recommended that Police use a rigorous process "to determine the appropriate cut-score on the various assessments it is using in the selection of Police Recruits" (2007, p. 8). In addition, Hattie (2007) considered the current use of stanines to screen applicants as unnecessarily risking misinterpretation and misapplication of GRT2 results.

## Personality Testing (15FQ+)

Police began using the 15 Factor Questionnaire Plus (15FQ+) in 2004. The 15FQ+ is a personality test developed by Psytech International for use in industrial and organisational settings (Tyler, 2003). It is designed to assess a range of personality dimensions (e.g. Affected by Feelings - Emotionally Stable, Accommodating - Dominant, Trusting - Suspicious, and Confident - Self-doubting). The 15FQ+ is considered capable of 'mapping' the 'big-five' personality dimensions also measured by the NEO-PI-R, the instrument previously used by Police. A number of the NEO-PI-R dimensions had displayed significant relationships with recruit training success (Black, 1995).

The 15FQ+ is administered by trained administrators who obtain a computer-generated report and then compare an applicant's scores to an 'Ideal' role profile. Where an applicant's scores differ by more than one sten from the 'Ideal' on a dimension, interview questions are identified from a provided list for use in the Behavioural Interview to confirm or refute any areas of concern (OPRA Consulting, 2004c). The 15FQ+ is not a pass/fail test but intended as a means to identify issues and concerns which can be further explored during the Behavioural Interview, "to flag issues that are followed up and if required can be referred to a psychologist" (New Zealand Police, 2004, p. 1).

### Meets standards

The reliability of the 15FQ+ was considered sound because:

- > Reliability levels described by Tyler (2003) and reported in the 15FQ+ technical manual (Psychometrics, 2002) are considered adequate for this test and its current application. Test-retest reliability coefficients reported in the technical manual fall within the levels expected and internal consistency levels are reasonable given the small number of items in some scales.

The validity of the 15FQ+ for the assessment of cognitive ability is supported by:

- > The validity of this test is comparable with other widely used and well-respected personality tests. The 15FQ+ "has frequently demonstrated exceptional construct internationally as well as good criterion-related validity in South Africa" (Tyler, 2003, p. 11).
- > A methodical and defensible tendering process was used to ensure the 15FQ+ was a sound measure of the key domains measured by the previously used NEO-PI-R personality inventory.

Assessment administration policies and guidelines are considered sound because:

- > Test administration instructions in the supporting documentation are clear, succinct and specific.
- > Supporting technical documentation is comprehensive and clearly described. Documentation included descriptions of instrument construction and specifications, interpretation strategies including impression management scales, equivalent forms, and psychometric properties including reliability, validity and normative data (Psychometrics, 2002). Additional information is also available on the publisher's web site (<http://www.psytech.co.uk/tests/15FQ+.htm>).

The potential discriminatory impact of the 15FQ+ is considered minimal, particularly given the manner in which it is used by Police. The evidence provided in the technical manual indicates that the internal consistency of the scales was broadly equivalent for men and women, and for respondents drawn from different ethnic backgrounds (Psychometrics, 2002).

### **Scope for improvement**

Validity information providing empirical evidence for the relationships between 15FQ+ factors and Constable performance is not yet available, indicating that interpretations based on the current 'ideal 15FQ+ profile' need to be treated with caution. While a structured process using Police subject matter experts created the ideal profile as an interim measure only (OPRA Consulting, 2004b), Police have not yet taken action to conduct empirical validation research. Sound empirical data describing the relationship between test scores and actual performance is a preferred source of evidence where interpretations of tests scores are involved (AERA, 1999), a fact recognised by the test provider, who has proposed a process for empirical validation (Barrett, 2005a). Police currently use the results of the 15FQ+ as a way of signalling potential concerns, i.e. to identify questions to be asked in the Behavioural Interview and during reference checking, and not to include or exclude applicants. However, it is essential that any assessment that may influence selection outcomes, even indirectly, is valid.

### **Physical Appraisal Test (PAT)**

The current form of the PAT (and of the Physical Competence Test, or PCT, which applicants complete later in the recruitment process) was established in 1993. These assessments were revised in response to a review, to ensure that Police Recruits had the physical ability and fitness to meet the needs of front-line policing and RNZPC firearms training, and to set standards that conformed with the requirements of the Human Rights Act 1993 (Wilson & Toomey, 1993). The standards in the PAT test were revised in 2006 to better reflect gender differences (Annan, 2006) without the nature of the test being changed.

The PAT assesses an applicant's physical ability and fitness. It consists of a 2.4 kilometre run to test aerobic fitness and endurance, a standing vertical jump test to assess leg strength, a push-up test to assess upper body strength and a grip strength test to assess grip and forearm strength. An applicant's run time is combined with a measure of their Body Mass Index (i.e. body weight in kg divided by height in metres squared) to give a rating between zero and six. Other test scores are converted to ratings between zero and three. All ratings are combined to give an overall score used to predict likely performance levels on the PCT. The PAT is not intended to be used as a pass/fail assessment, but if an applicant's ratings suggest they are likely to fail the PCT, they are advised to increase their fitness or strength (H. Macdonald, memorandum, 12 April, 2000).

### **Meets standards**

The validity of the PAT as an assessment of physical fitness is supported by the methodical process used to design its current form. Qualified human performance specialists conducted the design project (Wilson & Toomey, 1993), based on earlier research where the required fitness levels of Recruits was reviewed (Wilson & Toomey, 1990). The specialists sought input from the Police Medical Advisor, physical education specialists, and instructors with expertise in areas requiring physical ability levels (e.g. firearms and defensive tactics), and took into

account research findings, physical tests used for similar roles in other jurisdictions. and Otago University's Human Performance Centre Database information.

For the purposes that the PAT test is used, the minimum standards and classifications derived from assessment scores appear soundly based (Wilson & Toomey, 1993).

Administration instructions provided were clear and specific, supported by diagrams and written procedures. Forms are provided to assist collating and scoring of applicant performance.

### **Scope for improvement**

No studies describing the reliability of the PAT or the relationships different components of the PAT had with measures of performance were provided.

Administration guidelines reviewed were not consistent in how the role of the PAT is described. While the PAT was not designed to be used as a test that must be passed (or can be failed), references to the PAT in Police documentation suggest it is being used in some instances as a pass/fail test. In one instance, the PAT is described as an entry requirement, i.e. "If passing the PAT is to remain a requirement for entry" (Annan, 2006) and on the current PAT scoring forms, a 'PASS / FAIL' option follows each sub-test.

While Police have reviewed standards from time to time, responding to concerns over potential discrimination on the basis of gender and age and consistency with human rights legislation (Annan, 2006; Wilson & Toomey, 1990, 1993), no systematic process or recent review was identified. An example of an aspect of the assessment which could benefit a change in Constable requirements is the grip strength test, which was originally designed to meet the safe use of firearms no longer in current use. In addition, the recent changes to gender-based run times did not follow a sound and rigorous process to ensure the changes were defensible (C. Button, personal communication, 5 October, 2007).

### **Behavioural Interview**

The one and a half to two hour Behavioural Interview is conducted by the District Recruiting Officer using a standard interviewing format and questions drawn from a competency-based interview questions list (New Zealand Police, 2005b). The current interview format was introduced in 2005 under the guidance of a Registered Psychologist who was a Recruitment Manager employed by Police (G. Van Ooyen, personal communication, 18 September, 2007). The interview covers New Zealand Police values and competencies (i.e. Professionalism, Respect, Commitment to Maori and Treaty, Accountability for Performance, Building Partnerships, Challenging for Continuous Improvement, Exercising Judgement, Influential Leadership and Effective Communication). An applicant is also asked questions to explore areas which diverge significantly from an 'ideal score' on the 15FQ+ they completed. Interviewers rate applicant answers on a five-point rating scale (1 = no evidence of competence through to 5 = strong evidence of competence).

### **Meets standards**

Validity of the interview was assisted by:

- > Clear descriptions of the KSAOs assessed were included in the assessment and logical linkages were observed between the questions asked and these KSAOs.

- › The provision of questions developed by registered psychologists to assist interviewers to confirm or refute any areas of concern highlighted by an applicant's personality test results (OPRA Consulting, 2004c).

### **Scope for improvement**

Reliability of the interview process was limited by a lack of guidelines for conducting the interview and interpreting and rating responses.

Pass marks or standards expected of applicants were not defined.

Validity of the interview was limited by:

- › The process used to establish and justify the areas assessed and the standards against which applicants were evaluated were not described.
- › Mechanisms to enable a review of the validity of interview assessments were not described.

Administration guidelines provided insufficient guidance for the interviewing process (i.e. interviewing policy, administration requirements, interpretation and moderation procedures, rationale for rating or classification of interviewees, description of how information obtained was to be used, etc).

Mechanisms to enable regular review of the reliability and validity of the Behavioural Interview were not described in the information provided.

### **Physical Competence Test (PCT)**

The PCT assesses an applicant's physical ability and fitness and was established in 1993 following a review of physical fitness testing (Wilson & Toomey, 1993). A recent adjustment to a wall scaling task was made for applicants (Annan, 2006). The PCT consists of a 400 metre course involving a series of 12 physical tasks (e.g. pushing a trailer for 10 metres, carrying a wheel assembly for 10 metres, running 200 metres, running along a 5 metre 'L' beam, climbing through a window 1.3 metres off the ground, and dragging a 74 kilogram weight for 7.5 metres). The total time taken to complete the tasks is recorded and used to assign applicants a category score (Wilson & Toomey, 1993). Applicants who do not complete the PCT are not accepted for training.

### **Meets standards**

The reliability of the PCT was assisted by a clear documented description of standards required for passing (Physical Education and Defensive Tactics Section, 2004).

The validity of the PCT was assisted by a methodical process conducted by the qualified human performance specialists (Wilson & Toomey, 1993) being used to develop the PCT and an explicit description of how the content of the PCT related to a Constable's work (Physical Education and Defensive Tactics Section, 2004). Based on the documentation provided, descriptions of the relationships the PCT had with measures of performance included the finding of a significant correlation of .79 that the PCT displayed with a 12 minute Coopers run test (Wilson & Toomey, 1993).

Classifications derived from PCT assessment scores appear soundly based. The minimum entry standard of 3 minutes was primarily based on distribution of PCT scores for 566



Recruits graduating over 1992 and 1993 (Wilson & Toomey, 1993). However, a systematic process to review assessment processes and cut-scores has not been conducted in recent years.

Administration instructions were clear and specific, supported by diagrams, testing session guidelines, and forms to collate and score applicant performance.

Review processes are in place to help minimise the potential for injury during assessment. As described below, due to the incidence of PCT-related injuries and a desire to examine the operational relevance of PCT assessments, a research project is currently being considered (S. A. Bruce, personal communication, 28 September, 2007).

### **Scope for improvement**

Defensibility of current standards may be reduced, as a systematic process to review the accuracy of minimum standards required to pass the PCT does not appear to have been conducted. Given that the last comprehensive review was in 1993 and a range of changes regarding what is required of Constables could reasonably be expected, the relevance of current standards can be questioned. Recent changes to wall height standards lacked a sound rationale or research evidence on which to base the changes (C. Button, personal communication, 5 October, 2007).

The impact on Recruits of the current PCT has been questioned. Following a review due to the number of Recruits suffering injuries when taking part in a PCT, changes were made to how the PCT is administered which reduced injuries (S. A. Bruce, personal communication, 3 October, 2007). However, concerns over injuries still exist, signalling a need for the mechanisms used in the PCT to be reviewed, preferably against the physical demands actually required of Police Officers. Such a review may also offer opportunities for Police to increase the accuracy and efficiency of physical fitness testing, given the developments over the last fifteen years in knowledge about exercise physiology and the availability of a range of sophisticated testing devices.

### **First Aid Certificate**

Applicants are required to provide a First Aid Certificate showing that they have met the requirements of NZQA units 6400, 6401 and 6402. Applicants who cannot provide a Certificate are not accepted for training.

### **Meets standards**

The validity of this assessment is supported, in that the requirement for a First Aid Certificate is linked directly to operational requirements (i.e. to ensure Police Officers are able to provide first aid if required in the course of their duties).

The minimum standard that applicants must achieve is specific and clear for applicants and for District Recruitment Officers who are responsible for assessing whether applicants meet the standard (i.e. meets NZQA unit standards 6400, 6401 and 6402).

The administration policies and guidelines appear effective. The consistency with which assessments could be administered is assisted by the requirement that assessments are conducted by certified assessors. The information about requirements given to applicants is clear and specific, as evidenced in the documentation provided.

### **Scope for improvement**

In the information provided, mechanisms to enable regular review of the types of First Aid skills required were not described.

The process used to establish and justify the required standard of First Aid skills was not clear from the documentation provided.

### **Defensive Driving Certificate**

An applicant is required to provide a Defensive Driving Certificate showing that they have met the requirements of Land Transport New Zealand. Applicants who cannot provide a Certificate are not accepted for training.

### **Meets standards**

The validity of this assessment is supported, in that the requirement for a Defensive Driving Certificate is linked directly to operational requirements (i.e. to ensure Police Officers have a sound basic understanding of defensive driving to apply in the course of their duties).

The minimum standard that applicants must achieve is specific and clear for applicants and for District Recruitment Officers who are responsible for assessing whether applicants meet the standard (i.e. meets Land Transport New Zealand requirements).

### **Scope for improvement**

The process used to establish and justify the standard of defensive driving required was not clear from the documentation provided.

In the information provided, mechanisms to enable regular review of the necessity of the requirement were not described.

### **Typing and Computer Skills Test**

The Typing Test was introduced in 1994 and assesses an applicant's typing speed and accuracy (G. van Ooyen, personal communication, 14 September, 2007). Applicants complete a 10 minute timed test where they are required to sit at a computer, 'open' MS Word, create a blank document and type the text from a prescribed sheet. The test is typically administered by the Recruiting Officer before the Behavioural Interview. Applicants must achieve a speed of 25 accurately typed words per minute. Applicants not achieving the speed required are allowed to re-sit the test throughout the recruitment process until they do so. Applicants who do not successfully complete the test are not accepted for training.

### **Meets standards**

The validity of this assessment was supported, in that the need for assessment of typing and computer skills was linked to operational requirements (i.e. keyboard and word processing work required from Police Officers during the course of their duties, such as entering information into Police intelligence systems, completing reports and preparing documentation for court proceedings).

The minimum standard that applicants must achieve is specific and clear for applicants and for District Recruitment Officers who are responsible for assessing whether applicants meet the standard (i.e. 25 accurately typed words per minute).

### Scope for improvement

The process used to establish and justify the standards required for the Typing and Computer Skills Test were not clear from the documentation provided.

The administrative policies and guidelines were inadequate, in that limited instructions were provided to some administrators which would likely reduce the reliability of the tool.

In the information provided, no mechanisms to enable regular review of the reliability and validity of the Typing and Computer Skills Test were evident (i.e. it appears that the test in its current form has remained unchanged since its introduction 13 years ago).

### Reference Checks and Home Visit

Reference Checks on an applicant's background are made by a Recruitment Officer or Selection Specialist using a structured Referee Interview Guide (New Zealand Police, 2005b). Applicants are asked to nominate as referees past and present employers, people of good standing in the community and officials of any clubs they belong to. The checks cover the same New Zealand Police values and competencies as covered in the Behavioural Interview, use the same rating scale and involve a minimum of six different referees.

A Recruitment Officer or Selection Specialist will make a Home Visit to an applicant's home to assess their home environment, family stability and family support for their application (New Zealand Police, 1997a). Information on the applicant's home situation is collected using a structured 'Recruiting Enquiries Form' that does not require a rating or assessment.

This feedback from the Reference Checks and Home Visit is used as supporting information to assist the District Recruiting Officer make an overall selection recommendation.

### Meets standards

The focus of the Reference Checks was assisted by a description of the information sought during a Home Visit and a 'Recruiting Enquiries Guide'. A range of questions was provided for each to assist those conducting the Reference Check to identify questions to confirm or refute any areas of concern highlighted by an applicant's personality test results.

### Scope for improvement

The reliability of the Reference Checks and Home Visit was limited by:

- > Guidelines for conducting Reference Checks gave insufficient guidance to help ensure consistent processes were followed, and that interpretation and rating of the information gathered was reliable. The documentation lacked comprehensive descriptions of the procedures to be followed, criteria describing standards expected and rating procedures.
- > Guidelines for conducting Home Visits gave insufficient guidance for the interpretation and rating of the information gathered. While the information to be sought was described, criteria describing standards expected and rating procedures was lacking.

No evidence was provided of the process used to establish and justify the standards against which applicants were assessed using the Reference Check and Home Visit.

In the information provided, mechanisms to enable review of the validity of Reference and Home Visit assessments were not described.

## SCOPE

SCOPE is a realistic job preview exercise that requires an applicant to complete 40 hours of practical field observation and a pre-Recruit Course reading exercise (Research and Development Group, 1999; New Zealand Police, 1997b). The field observation takes place at a police station or on patrol to enable an applicant to experience a wide range of police work first-hand. Applicants are required to observe and investigate different facets of the Police as an organisation and the work of Police Officers, and record their findings in a SCOPE Workbook. They are guided by an assigned Supervising Officer who supports the applicant and assesses their behaviour during the exercise (e.g. using an assessment form covering appearance, punctuality, communication, temperament, etc).

The Pre-Course Reading Test requires an applicant to read prepared information on the New Zealand Police, the New Zealand legal and judicial system, and New Zealand Police ethics and values. On completion of the reading, the applicant sits a 28 question multi-choice test on the content of the material and is required to attain a score of 20 correct answers. Applicants not achieving a pass mark are given the opportunity to re-sit the test until they pass.

The assessments and feedback from SCOPE assessments, and how the applicant found the experience, are discussed with them. The information arising from the SCOPE exercise and the applicant's comments is used as supporting information to assist the District Recruiting Officer make an overall selection recommendation.

### Meets standards

The reliability of the Pre-Course Reading Test is assisted by clear instructions for administering and interpreting the test.

Administration information provided some guidance to applicants regarding their responsibilities, expectations of applicants, exercise directions and resource information.

### Scope for improvement

Standards of performance, classifications or pass marks for the SCOPE Workbook exercise and the assessment of the applicant during the SCOPE exercise were not clearly described in the documentation provided.

Administration guidelines provided insufficient guidance to workplace supervisors to ensure reliable evaluation of an applicant's SCOPE Workbook and behaviour while completing SCOPE (e.g. lack of assessment guidelines, criteria against which ratings were to be made, etc).

Mechanisms to enable the reliability and validity of SCOPE assessments to be reviewed were not described.

## Final Medical Clearance

The Final Medical Clearance is the second stage in the assessment of an applicant's medical fitness (New Zealand Police, June, 2007a). Applicants who have successfully passed all recruitment assessments undergo a medical examination by Police accredited practitioners (i.e. ophthalmic and registered medical practitioners). Areas assessed include cardiovascular, respiratory, ear, nose and throat, hearing, vision, digestive, genito urinary, metabolic, nutrition, endocrine, neurological, orthopaedic, cancer, diabetes, obesity, and mental and psychological health. The process involves applicants completing a 'Police Applicant Medical

Questionnaire Form' and then being examined and tested by relevant practitioners who complete a report on their findings (e.g. Visual Examination Report, Ophthalmological Certificate for applicants who have undergone refractive surgery, Police Medical Examination Report and Recruit Applicant Asthma Questionnaire). Applicants who do not meet the medical requirements are not accepted for training. If an applicant is deemed to have met Police medical standards, they are able to complete a PCT as the final stage in their recruitment process.

### **Meets standards**

The reliability of assessment processes appears appropriately supported by clear procedural and reporting guidelines (i.e. examination guidelines and protocols and forms for certified specialist reports).

The Final Medical Screening Clearance standards appear soundly based in that they were set using a methodical process, based on a range of credible sources (e.g. specialist ophthalmic advice, procedures and standards used by comparable agencies such as the New Zealand Defence Force, the New Zealand Fire Service and Police Services in the United Kingdom, and medical research), and ensured recommendations made complied with statutory requirements (e.g. Land Transport Act 1998) relevant to Police (Robinson, 2006).

It can be reasonably expected that having qualified medical practitioners conduct the assessment using standard diagnostic procedure to evaluate the risk factors involved, will help ensure that the Medical Screening Clearance validly measures those aspects it is designed to do so.

### **Scope for improvement**

In the information provided there appears to be an absence of mechanisms to enable regular review of the reliability and validity of medical standards (Robinson, 2006). Robinson considered that it was important for Police to record instances of potential Recruits being found medically unfit so the "information would also act as an audit to ensure consistency in the application of medical standards" and to keep a record of Officers who medically retire so that "a profile of medical disengagement can be aligned to entry standards" (2006, p. 6).

## **Recruit training assessment**

Recruit performance on the 19 week Recruit Training Course is assessed using eight separate assessments (i.e. three Written and two Practical Summative Assessments, and three Operational Assessments). Below, each assessment method used is described along with the KSAOs it has been identified to assess, how assessment results are used and how the assessment method compares with the baseline standards.

### **Written Summative Assessments**

The Written Summative Assessments are two to three hour written examinations, each covering a specific area. Recruits are required to achieve a pass mark of 60% in each.

### **Summative Written Examination 1**

A written examination covering ethics, Maori responsiveness, quality customer service, the law, policies and practical procedures relevant to information gathering, report documentation, and incident and offence reporting. Summative Written Examination 1 is held around week 4 of the Recruit Course.

### **Summative Written Examination 2**

A written examination covering the law, policies and practical procedures relevant to dealing with suspects, response offences and witnesses, and the judicial setting. Summative Written Examination 2 is held around week 8 of the Recruit Course.

### **Summative Written Examination 3**

A written examination covering the law, policies and practical procedures relevant to dealing with domestic violence, road safety, controlled drugs, firearms, psychological patients, anti-social groups and youth, sexual offences, alcohol and intoxicants suspects, and further aspects of response offences and witnesses and the judicial setting. Summative Written Examination 3 is held around week 18 of the Recruit Course.

The procedures followed for Recruits who fail to achieve a pass mark (60%) on the written summative assessments are as follows:

- > Recruits who fail the first and/or second Summative Written Examination are placed on a 'First Warning', and then a 'Second Warning' if they fail a second time. A Performance Management Plan is developed for Recruits on warning (including the provision of additional tutoring). These Recruits are, however, not required to re-sit these assessments, and are permitted to continue with the course. If they pass the third Summative Written Examination, they graduate normally.
- > If a recruit fails the third Summative Written Examination, they are given up to two opportunities to re-sit the exam. This may necessitate being placed on Special Leave for further study, which would prevent graduation with the recruit's Wing (i.e. with their classmates). If at any stage a recruit accumulates three 'fails', their final examination paper is remarked. If the remarking supports the fail decision, following consultation, discussion and seeking of an explanation the RNZPC may seek termination of employment.

### **Meets standards**

Reliability of these assessments was considered sound because:

- > Clear and specific exam instructions are given to Recruits (e.g. the focus of the examination, resources needed and how to respond to questions).
- > Comprehensive marking guidelines are provided to instructors responsible for assessment, including provision of model answers.
- > Instructors independent of markers are used to moderate assessments to help ensure consistency of marking standards.
- > An opportunity is provided for Recruits to review their own marked responses and scores and if they feel a mistake has been made, they can apply for a review (with the

support of their Sergeant). Recruits also have the opportunity to provide feedback to instructors on exam questions.

- › Post-examination procedures are used to ensure changes and corrections to the 'exam bank' are consistently made.

Validity of the assessment is supported through use of a methodical quality assurance process to produce each assessment. For example, the instructor responsible for the assessment verifies the topics covered prior to the exam with the Recruit Wing involved, uses the RNZPC Exam Bank to compile the exam, and passes the exam to a senior Police Studies Officer for review before finalising and producing the exam.

Administration policies and guidelines were considered clear and complete because:

- › Assessor guidelines include quality control checklists for briefing, conducting examinations, debriefing markers, explicit descriptions of marker responsibilities, and guidelines for marking and moderation.
- › Recruit information is comprehensive and includes the objective of each assessment, the opportunity to complete a formative assessment two to three weeks before a summative assessment, reassessment procedures, feedback policies and review processes (e.g. Initial Training Group, 2007a).

Review policies and procedures were considered sound. RNZPC procedures require 'signed off' completion of a quality review process accompanying development and use of every Summative Written Examination.

### **Scope for improvement**

Validity of assessment tool. While feedback from operational staff on recruit performance (i.e. as Probationary Constables) and preparedness for operational work is sought, no evidence was provided that the relationship between Summative Written Examinations and measures of job performance has been examined.

Justification for pass mark level. No evidence was found justifying the rationale for examination pass levels being set at 60%. While comments made by senior instructors indicate that care is taken to ensure Recruits demonstrate an acceptable level of knowledge on 'essential' topic elements, how this was consistently achieved was not documented. Using a soundly-based pass mark is important in this context given that Recruits are able to fail Written Summative Examinations One and Two and still graduate by passing Written Summative Examination Three. In such a case, it is not clear how the current pass mark would ensure that the recruit has demonstrated the range of knowledge expected of a RNZPC graduate.

### **Practical Summative Assessments**

In addition to the Written Summative Assessments, recruit performance is assessed in two Practical Summative Assessments. Recruits are required to achieve a pass mark of 80% in each. Practical Summative Assessment 1 is held approximately in week 6 of the course and Practical Summative Assessment 2 is held approximately in week 17. Both practicals consist of a scenario requiring each recruit to respond to and investigate a complaint regarding a possible offence (e.g. reported stolen vehicle). Recruits are briefed on the scenario, deal with complainant(s) who are role-played by RNZPC staff and then submit relevant reports,

statements and records resulting from their response and investigation. Recruits are assessed on how they deal with the complaint and the information they submit.

The procedures followed for Recruits who fail the Practical Summative Assessments are identical for those used with the Written Summative Assessments.

### **Meets standards**

Reliability of these assessments was considered sound because:

- › Recruits are given comprehensive scenario instructions (e.g. checklist-based verbal and written explanations of assessment objective, procedures, resources and expected responses).
- › Rehearsal and pre-assessment of the Practical Examination involving role-players and examiners are used prior to conducting the scenario.
- › Instructors independent of markers are used to moderate assessments to help ensure consistency of marking standards.
- › Assessment procedures incorporate standard criteria, anchored rating scales and a standardised, computer-based rating integration process to produce examination ratings.
- › An opportunity is provided for Recruits to review their own marked responses and scores, and to provide feedback to instructors on exam questions. If Recruits feel a mistake has been made, they can apply for a review (with the support of their Sergeant).

Validity of the assessment is supported through use of a methodical process to confirm that the assessment is accurately focused on the relevant KSAOs. There is a documented process for compiling and reviewing each practical assessment. The assessment scenario is drawn from a number of standard scenarios, modified where required and reviewed by independent instructors using a quality review process (i.e. pre and post scenario use) which meets 11552 and 4098 Unit Standard requirements.

A reasoned process has been followed to help ensure pass marks and KSAOs assessed are equivalent across assessments. A structured process led by a registered psychologist and involving a range of individual subject matter experts was used to develop an assessment procedure for both Practical Summative Assessments. The assessment procedure differentiates those criteria critical to competence and those that are not. Criteria considered critical require demonstration of a 'mastery' level of performance to achieve any marks for that criterion. Criteria not considered essential accrue marks for different levels of performance (e.g. on a three-point scale). Where a recruit fails on critical criteria, they are required to complete remedial learning until mastery has been demonstrated.

Administration policies and guidelines were considered clear and complete because:

- › Assessor guidelines include quality control checklists for briefing, conducting examinations, debriefing role-players and markers, descriptions of assessment team and recruit responsibilities, occupational health and safety requirements, and guidelines for marking and moderation.
- › Recruit information is comprehensive and includes the objective of each assessment, the opportunity to complete a formative assessment two to three weeks before a



summative assessment, reassessment procedures, feedback policies and review processes (e.g. Initial Training Group, 2007a).

Review policies and procedures were considered sound. RNZPC procedures require 'signed off' completion of a quality review process (i.e. to 11552 and 4098 Unit Standard level) following use of a scenario for a practical assessment.

### Scope for improvement

Validity of assessment tools. While feedback from operational staff on recruit performance (i.e. as Probationary Constables) is sought, no evidence was provided of the relationship between Practical Summative Examinations and measures of performance (i.e. of Probationary Constables and Constables).

Justification for pass mark level. Although a systematic and structured process was used to help ensure consistency across different practical examinations, the rationale for the pass mark being set at 80% was not evident in documentation.

### Driver Training

Held at week 9 over nine days and includes driving, basic crash investigation and Police patrolling including urgent duty response and pursuit driving (Professional Police Driver Programme, 2005; Professional Police Driver Programme, 2007). Recruits must attain a 'silver' classification, as defined by the Professional Police Driving Programme (PPDP), to graduate. Recruits must attain 100% pass mark with a 'standard' Road Code Test, a 100% pass mark on the written driving assessment and achieve competency with the various driving skills. Tutorial assistance is provided to those who do not initially attain the required competencies. If following tutorial assistance the recruit is unable to pass the assessment, the recruit will not be permitted to graduate.

### Meets standards

The reliability of these assessments was considered sound because:

- > Descriptions of the assessment objectives, areas covered and standards expected are given to Recruits prior to assessment.
- > Comprehensive guidelines are provided to instructors responsible for assessment, including criteria, standards of assessment, moderation processes, etc. (Professional Police Driver Programme, 2007).
- > Recruits have the opportunity to request a review of their assessment.

Validity of the assessment is supported through:

- > Consistency of assessment standards with legislative requirements (e.g. Land Transport Act 1998), Police General Instructions on pursuits, urgent duty driving and use of Police vehicles, and Police policies.
- > Pass marks and KSAOs assessed are based soundly on Professional Police Driver Programme requirements (Professional Police Driver Programme, 2005).

Administration policies and guidelines were considered clear and complete because:

- › Assessor guidelines include quality control checklists for briefing, conducting assessments, responsibilities of the recruit, safety requirements, and guidelines for reassessment, marking and moderation.
- › Recruit information is comprehensive and includes the objective of the assessment, assessment standards required, reassessment procedures, feedback policies and review processes.

Review policies and procedures were considered sound. RNZPC procedures require 'signed off' completion of a quality review process (i.e. to 11552 and 4098 Unit Standard level) following an assessment.

### **Scope for improvement**

Validity of the Driver Training Assessment does not appear to have been comprehensively examined against road policing role requirements. A research project has been initiated to evaluate the relationship between Probationary Constables' and Constables' driving incidents and the assessments of the driving performance of those same Officers when they were at RNZPC. This research will assist in identifying changes needed in some aspects of the training and assessment of Driver Training (S. A. Bruce, personal communication, 28 September, 2007).

### **Firearms Training**

Held at approximately week 11 over eight days. Recruits must attain competency (i.e. 100%) in the safe operational use of the Police Glock handgun and Bushmaster rifle (i.e. competence in all tests of weapon handling skills as per Test of Elementary Training Skills (TOETS)) and pass a written examination. The Glock qualification shoot minimum mark is 19 out of 24. The Bushmaster qualification shoot minimum mark is 21 out of 24. The pass mark for the written examination is 80%. Tutorial assistance is provided to those who do not initially attain the required competencies. If following tutorial assistance the recruit is unable to pass the assessment, the recruit will not be permitted to graduate.

### **Meets standards**

The reliability of the Firearms Training assessment was considered sound. Guidelines for instructors responsible for assessment were specific (e.g. Staff Safety Tactical Training, 2005a and Firearms Training Section, 2005).

Validity of the assessment is supported through:

- › Consistency of assessment standards with legislative requirements (e.g. Crimes Act 1961), Police General Instructions (e.g. F186, F187 and F188) and Police policies relevant to firearms use (e.g. Safety in Training Policy).
- › Involvement of Police firearms specialists in the development of operator guidelines and assessment standards and the requirement that assessment standards maintain consistency with operational requirements as set by New Zealand Police Operations Group directives (M. Stonyer, personal communication, 28 September, 2007).

Pass marks and KSAOs assessed are based on the risks involved in use of firearms and Police operational requirements (S. A. Bruce, personal communication, 28 September, 2007),

although no formalised documentation currently exists (M. Stonyer, personal communication, 28 September, 2007).

Administration policies and guidelines were considered clear and reasonable for certified firearms instructors. Administration guidelines included a description of the assessment process and the individual assessment standards for each 'test' making up the assessment.

Review policies and procedures were considered sound. Firearms training and assessment standards are required to maintain consistency with New Zealand Police Operations Group directives (M. Stonyer, personal communication, 28 September, 2007).

### **Scope for improvement**

Mechanisms for using participant feedback to ensure assessment processes were considered fair and reasonable were not documented.

### **Defensive Tactics**

The Defensive Tactics assessment determines competence in staff safety and tactical options including correct use of handcuffs, oleoresin capsicum spray and baton (side-handled and extendable), and the use of the carotid hold (Staff Safety Tactical Training, 2005b; Staff Safety Tactical Training, 2003). With the exception of the side-handled baton, a 60% pass mark is required for each of the Defensive Tactics skills. Tutorial assistance is provided to those who do not initially attain the required competencies. If following tutorial assistance the recruit is unable to pass the assessment, the recruit will not be permitted to graduate.

### **Meets standards**

Reliability of these assessments was considered sound because:

- › Descriptions of the assessment objectives, areas covered and standards expected are given to Recruits prior to assessment.
- › Comprehensive guidelines are provided to instructors responsible for assessment, including criteria, standards of assessment, etc.
- › An opportunity is provided for Recruits to review their assessments and if they feel a mistake has been made they can apply for a review (with the support of their Sergeant). Recruits also have the opportunity to provide feedback to instructors on assessments.
- › Assessor guidelines include checklists for briefing, conducting assessments, and guidelines for marking and moderation.

Validity of the assessment is supported through:

- › Consistency of assessment standards with legislative requirements (e.g. Crimes Act 1961), Police General Instructions on use of force, carotid hold, etc., and Police policies.
- › A research programme has been implemented to evaluate the use of force by recently graduated Probationary Constables, to identify the training that was most relevant to the action taken and to help ensure that training and assessment strategies are relevant to current operational conditions (K. J. Bruce, personal communication, 28 September, 2007).

Administration policies and guidelines were considered clear and complete because:

- > Assessor guidelines include checklists for briefing, conducting assessments, and guidelines for marking.
- > Post-examination moderation procedures are used to ensure changes and corrections to the 'exam bank' are consistently made:
- > Recruit information is comprehensive and includes the objective of the assessment, assessment standards required, reassessment procedures and feedback policies.

Review policies and procedures were considered sound. RNZPC procedures require 'signed off' completion of a quality review process (i.e. to 11552 and 4098 Unit Standard level) following an assessment.

### Scope for improvement

Justification for pass mark level. The rationale for the pass mark being set at 60% was not evident in documentation provided.

## Workplace Assessment Programme

Probationary Constables are required to demonstrate competence on the following ten Probationary Constable Workplace Assessment Programme (PCWAP) Standards within two years of graduation and complete LEGL 114, a Stage 1 Victoria University paper.

- > **Standard 1: Manage incidents and offences**  
Includes the management, investigation and documentation of incidents and offences.
- > **Standard 2: Report and record incidents and offences**  
Includes accurately reporting and recording incidents and offences, and demonstrating an understanding of the nature of the offence and the related law.
- > **Standard 3: Conduct interviews**  
Includes demonstrating an understanding of the role of interviewing and the ability to conduct planned and lawful interviews.
- > **Standard 4: Perform road policing duties**  
Includes demonstrating an understanding of the purpose of and the ability to perform road policing duties.
- > **Standard 5: Carry out patrol duties**  
Includes demonstrating knowledge of legislative powers and employing effective patrol techniques to detect and apprehend offenders, gather information and maintain police visibility.
- > **Standard 6: Carry out administrative processes**  
Includes demonstrating the ability to carry out administrative processes in the policing context.
- > **Standard 7: Manage incidents involving children and young persons**  
Includes demonstrating knowledge and skills to manage situations involving children and young persons within the care and protection of youth justice contexts.

› **Standard 8: Manage incidents involving mental health consumers**

Includes demonstrating the knowledge and skills to manage incidents involving mental health consumers.

› **Standard 9: Manage incidents of family violence**

Includes demonstrating the ability to manage and investigate incidents of family violence.

› **Standard 10: Use the Tactical Options Framework**

Includes demonstrating the ability to assess situations and use appropriate tactical options.

For each of the Standards, a Probationary Constable compiles evidence and demonstrates their competence to a Designated Assessor. Where one of the Standards is not met, two further reassessments of that Standard are permitted before failure to meet the Standard results in the implementation of a performance management process.

### Meets standards

The reliability of these assessments was considered sound because:

- › Clear and specific assessment instructions are provided to Probationary Constables, including a description of the focus of the programme, self-management strategies, assessment processes, support resources, reassessment and review procedures, and roles and responsibilities of those involved in the programme (Recruit Training Group, 2006).
- › Comprehensive guidelines are provided to assessors responsible for assessment, including descriptions of the roles of those involved, assessment procedures, criteria and reporting formats (Recruit Training Group, 2005a).
- › An opportunity is provided for Probationary Constables to review and/or appeal their assessments if they do not agree with the results and to provide feedback to the Recruit Training Group on the programme.

Validity of the assessment is supported through documented links between the assessment standards and relevant legislation (e.g. Children, Young Persons and Their Families Act 1998 and Domestic Violence Act 1995), operational procedures, General Instructions and Police policies (Recruit Training Group, 2007). Documentation links the assessment content to Police values, the Police focus on Treaty Principles and Maori values, and operational and legislative requirements (Recruit Training Group, 2006).

Administration policies and guidelines were considered clear and complete because:

- › Assessor guidelines include detailed descriptions of assessment procedures, question templates, interpretation of evidence gathered, providing feedback, reporting assessments, and procedures for moderation and review.
- › Probationary Constables' information is comprehensive and includes the objective of each assessment, assessment standards and performance criteria, reassessment and review procedures, and roles and responsibilities of those involved in the assessment process (e.g. Recruit Training Group, 2006).

Justification for pass mark level. The use of a '100%' pass mark is based on the rationale that all the requirements for each performance standard are necessary for a Constable to be effective in operational settings.

Review policies and procedures were considered sound because RNZPC conducts an annual review of assessment policies and processes at a National Workplace Assessors Conference (Recruit Training Group, 2005b).

### **Scope for improvement**

While feedback is sought from operational staff on the performance of recent graduates of the RNZPC in general, evidence that the relationship between the Probationary Constable Workplace Assessment Programme assessment standards and measures of job performance (i.e. of Constables) has been examined, was not obvious from the evidence provided.

### **LEGL 114: Introduction to criminal law and problem solving**

LEGL 114 is a Stage 1 university paper taught by Victoria University of Wellington. The assessment examines a Probationary Constable's understanding of the role of criminal law and criminal procedure, the operation and limits of the criminal justice system in resolving social problems, and their understanding of and ability to apply problem solving methodology in common law enforcement. Assessment is by university staff on the basis of four assignments, each meeting specific marking criteria. Probationary Constables must achieve a pass mark of 50%.

### **Meets standards**

Reliability of these assessments was considered sound because:

- › Clear and specific assessment instructions are provided to Probationary Constables, including a description of the focus of the course, assessment processes and standards (e.g. model answers) and support resources (Victoria Police Education Programme, 2007).
- › Tutors assessing student assignments are required to follow comprehensive marking and moderation guidelines (University Teaching Development Centre, 2004). Marking guidance is provided to tutors before the course begins and prior to each assignment, in addition to the Course Co-ordinator being responsible for reviewing a range of assignments marked by each tutor (J. Ramshaw, personal communication, 28 September, 2007).

Administration policies and guidelines were considered reasonable. Tutors were required to follow Victoria University's VPEP Quality Assurance Plan (J. Ramshaw, personal communication, 26 September, 2007) and guidelines for participating Probationary Constables' information was comprehensive (Victoria Police Education Programme, 2007).

### **Scope for improvement**

Although an obvious link exists between the need for Police Officers to effectively use criminal law and the content of LEGL 114, evidence for the range of KSAOs required and the contribution of the course to those requirements was not obvious from the evidence provided.

Justification for pass mark level. Although a 50% pass mark is consistent with common tertiary education practice, the rationale for the pass mark being set at this level was not evident in documentation provided.

In the information provided, mechanisms to enable review of the relationship between the requirements of the Constable role and the standards required by LEGL 114 assessments were not described.

## Summary evaluation of assessment methods

The levels of consistency with the second baseline standard is summarised in the table below.

Recruitment Assessment Tools	High	Reasonable	Low
Health Questionnaire		■	
Offences Clearance Check	■		
Swimming Certificate of Competency		■	
Cognitive Ability Testing (GRT2)		■	
Personality Testing (15FQ+)		■	
Physical Appraisal Test (PAT)		■	
Behavioural Interview			■
Physical Competence Test (PCT)		■	
First Aid Certificate		■	
Defensive Driving Certificate		■	
Typing and Computer Skills Test			■
Reference Checks and Home Visit			■
SCOPE			■
Final Medical Clearance	■		
Recruit Training Assessment Tools	High	Reasonable	Low
Written Summative Assessments		■	
Practical Summative Assessments	■		
Driver Training	■		
Firearms Training	■		
Defensive Tactics	■		
Workplace Programme Assessment Tools	High	Reasonable	Low
PCWAP Standards	■		
LEGL 114		■	

### Scale

<b>High =</b>	Close consistency between the assessment method and the baseline standard, indicating that a method is likely to assist Police select recruits with the targeted KSAOs.
<b>Reasonable =</b>	Reasonable consistency between the assessment method and the baseline standard, indicating that the method is likely to assist Police select recruits with the targeted KSAOs, but that there are specific opportunities to enhance its effectiveness.
<b>Low =</b>	A low level of consistency between the assessment method and the baseline standard, indicating that Police cannot be confident that the method will assist the selection of recruits with the targeted KSAOs.

An overall judgement of how well each assessment method was considered to meet the baseline standard was made based on the proportions of 'strong points' and 'scope for improvement' points, the impact these points were likely to have on the accuracy of assessment, and the nature of the instrument (e.g. the rigour expected of the administration, scoring and interpretation of a psychometric test was considered higher than that expected of a Swimming Certificate). These judgements are not definitive categorisations of the assessment tools, but judgements designed to assist the reader gain an overall picture of the conclusions.



As seen in the above table, the assessment tools used to assess Recruits tended to be less consistent with the criteria under this baseline standard than the tools used to assess Recruits during training and Probationary Constables during the Workplace Programme.

The key themes arising from the evaluation of how assessment tools were identified and applied are as follows:

- › The absence of a systematic and robust job analysis process prevents a soundly based assessment that the range of assessment methods adequately covers the KSAOs required.
- › Policies and processes to ensure regular review and responsiveness to changes in organisational requirements are largely not in place. For example, for ten of the fourteen assessment tools used during recruitment no evidence of review policies or processes was found, so reducing the ability of Police to ensure assessment processes remain effective over time.
- › The processes used to determine four of the fourteen assessment tools used to assess applicants met no criteria under this standard in full. The Behavioural Interview, Typing and Computer Skills Test, Reference Checks and Home Visit, and the SCOPE exercise were all of uncertain reliability and validity. In addition, these assessment tools are administered by a range of staff and rely heavily on the judgments of assessors but lack administration guidelines and soundly-based pass marks or a description of what is an acceptable standard. The ability of these tools to consistently provide accurate assessment results is uncertain.
- › The assessment tools used to assess Recruits during training and Probationary Constables are considered to be administratively sound, reliable, accurate, and supported by policies for regular review and improvement. While some gaps were identified in the rationale supporting pass marks and descriptions of links with measures of future performance, these tools can be considered capable of providing useful information to the assessment of Recruits and Probationary Constables.

## 8. Consistency of Practice

This section describes and reviews the evidence related to the third Baseline Standard.

### **Comprehensive and sound procedures have been used to ensure consistent application of assessment methods.**

The four criteria specifying what are considered comprehensive and sound procedures are detailed in the section titled '4. Baseline Practice Standards'. These procedures need to ensure that:

- › Individuals are appropriately qualified and competent in the use of the assessment tools for which they are responsible;
- › The integrity of assessment materials and confidentiality of assessment results are protected;
- › Reassessment policies and procedures are based on a sound rationale; and
- › Mechanisms are in place to ensure that administration, scoring, interpretation and communication of results is nationally consistent and matches recognised standards of assessment practice.

Following are the findings derived from documentation provided by Police, and discussions with Police National Headquarters Human Resources staff, District Recruitment Officers, Physical Education Officers, and Royal New Zealand Police College managers and instructors.

### Recruitment assessment

#### **Health Questionnaire**

##### **Meets standards**

Guidelines incorporated in the Health Questionnaire for applicants were comprehensive and clear (New Zealand Police, 2007a).

Applicants are provided with clear descriptions of the Questionnaire's purpose, the links between medical standards and job requirements and how the information is used (e.g. confidentiality and access).

All health questionnaires are evaluated nationally by a single individual who has experience working in the Police Medical Services Team and who has access to a specialist registered medical practitioner for advice as required.

#### **Offences Clearance Check**

##### **Meets standards**

The District Recruitment Officers who review applicants' offences declarations have access to assistance from experienced Officers and specialist intelligence units.

Consistency of application is supported by the requirement that marginal conviction clearances are submitted to and ruled on by the General Manager Human Resources. In addition, policy currently under discussion will require any discretion applied to an offences decision to be documented in a centralised Recruit Applicants Discretion Register (Annan, 2007a).

### **Scope for improvement**

An apparent lack of use of common interpretation guidelines. While Auckland District has developed guidelines to assist interpretation of the Policy on Proven Charges (New Zealand Police, 2007c) and guidelines have been developed by the National Human Resources team, comments made by District Recruitment Officers suggest that these guidelines are not commonly used, a practice which is likely to reduce the consistency with which Offences Clearance Checks are assessed.

## **Swimming Certificate of Competency**

### **Meets standards**

Specific assessor administration instructions are provided.

Assessors have been certified by a Police Physical Education Officer (PEO) using a standardised certification process (G. van Ooyen, personal communication, 18 September, 2007).

Designated individuals are responsible for monitoring and reviewing the standards and practices of the certified assessors awarding competency certificates (G. Wallwork, personal communication, 28 September, 2007).

### **Scope for improvement**

Comments by District Recruitment Officers suggested that in a small number of locations, swimming assessors do not always have access to pools of a sufficient depth for the test, so reducing the comparability of some aspects of the swimming assessment.

## **Cognitive Ability Testing (GRT2)**

### **Meets standards**

Procedures for training and qualifying Police staff responsible for administering and interpreting GRT2 test results were considered comprehensive and designed to ensure user qualifications are consistent with standards commonly expected for use of 'B' level tests in New Zealand (New Zealand Council for Educational Research, 2006). Training programmes were conducted by registered psychologists and included key psychometric and statistical principles, test user responsibilities, procedures, reporting of results and feedback strategies. Participants were required to successfully complete an open book examination before being certified to administer and/or interpret the GRT2.

Policies requiring confidentiality of GRT2 results and security of test materials are clearly documented. Testing policies require credentialed test users to take personal responsibility for the security of materials, ensuring unauthorised personnel do not have access to materials and that defaced or damaged materials are replaced.

Administration guidelines and instructions for decisions are clearly described, although as described below, there is likely to be some confusion created by different versions of these guidelines being in circulation.

### **Scope for improvement**

Different versions of the guidelines for deciding whether to 'stand down' or progress applicants appear to be documented in different forms, so risking inconsistent application of minimum standards (Hattie, 2007).

Retesting policies may be being applied inconsistently. While the average delay between testing and retesting was 63 days (well in excess of the minimum six weeks), seven applicants were retested without the required six week delay, and two were retested on the same day (Hattie, 2007).

Hattie's (2007) study found unacceptably high error rates in recorded test scores, poor recording practices and data-sets that did not correspond with each other. For Police to ensure practices are consistent with the baseline standards, Hattie recommended a review of current data management practices with a view to creating a dependable data-set of assessment scores that is useful to those who need access to the information.

## **Personality Testing (15FQ+)**

### **Meets standards**

Policies requiring confidentiality of 15FQ+ results and security of test materials are clearly documented. Testing policies require credentialed test users to take personal responsibility for security of materials, ensuring unauthorised personnel do not have access to materials and that defaced or damaged materials are replaced.

Administration guidelines and instructions for use of interpreted test results to identify questions for use in the Behavioural Interview and Reference Checks are specific and clear (OPRA Consulting, 2004c).

Computerised interpretation and scoring of 15FQ+ tests for all applicants ensures consistency.

### **Scope for improvement**

Only some Police staff using the 15FQ+ have the formal qualifications typically required for users of this test in New Zealand. The 15FQ+, as a personality test, is classified by the New Zealand Council for Educational Research (2006) as a Level C test, which requires users to have successfully completed advanced courses in the use of tests. Appropriate qualifications include in-depth 3rd year or postgraduate university courses in areas such as clinical psychology, personality theory and assessment. A number of test users do have the requisite qualifications, for example, at least five staff in the Auckland Police Recruitment team have graduate or post-graduate qualifications in psychology and psychometric assessment (W. Kennedy, personal communication, 26 September, 2007). However, others have no formal training in psychology or statistics beyond the four day course in the use of the 15FQ+. Individuals who are not qualified in testing and assessment may not know when guidance or support is necessary, as they may not fully understand the nature of the instruments and interpretations being made, or appreciate the ethical responsibilities they have as test users (Australian Psychological Society, 2002). These users may also find it difficult to respond

appropriately to unforeseen interruptions in testing, or to deal appropriately with the unexpected reactions of applicants receiving negative feedback on their test results. While the test provider provides specialist support and guidance, the provision of this support, especially to those in Districts where there are no suitably qualified staff, relies on users making a request. No framework or policy was observed that specified how qualified staff employed by Police take responsibility for supervising and providing support to unqualified users (e.g. for interpretation of results, identification of questions of interest and provision of feedback).

### **Physical Appraisal Test (PAT)**

#### **Meets standards**

Police Physical Education Officers (PEOs) are certified and monitored by senior PEO instructors to ensure administration, scoring and interpretation consistently follows standardised procedures in some Districts (G. Wallwork, personal communication, 27 September, 2007).

Administration instructions are specific and clear.

Resitting policies allow the PEO conducting the assessment to determine an appropriate amount of time between an unsatisfactory PAT and retesting (G. Wallwork, personal communication, 1 October, 2007). While no specific guidelines are documented, the policy is designed to enable a PEO to tailor the retesting time to best suit the individual's situation (e.g. level of fitness, gap between the result achieved and the desired standard, etc).

#### **Scope for improvement**

A mechanism to nationally monitor and evaluate PAT use was not described in the documentation provided (e.g. national monitoring and review procedures or a designated individual[s] with national co-ordination responsibilities). Comments made by District Recruitment Officers and PEOs suggested that practice in different Districts may vary, thus reducing the consistency with which PATs are conducted nationally. For example, while it is expected that Police Physical Education Officers (PEOs) carry out PAT assessments, the range of locations and limited availability of PEOs can result in PAT assessments being carried out by other Police staff (e.g. District Recruiting Officers).

The range of environments within which the PAT run is administered (i.e. city streets, a recreational park and an indoor gymnasium) can reduce the consistency of the assessment. For example, some Recruits being tested may need to avoid people and traffic, or run courses with tight corners, and others would not face the same conditions.

### **Behavioural Interview**

#### **Meets standards**

All recruitment staff conducting the interview have completed interview training as part of their accreditation to use the 15FQ+ (G. van Ooyen, personal communication, 18 September, 2007).

### **Scope for improvement**

Interviewing practices nationally appear to be variable. Comments made by District Recruitment Officers suggested that a range of approaches to interviewing is taken, in some cases using different interview guides and in other cases focusing solely on the issues identified in the 15FQ+.

No mechanism to promote national consistency of interviewing assessment was described in the documentation provided (e.g. monitoring and review procedures or a designated individual[s] with national co-ordination responsibilities).

No guidelines or policies governing reassessment using the Behavioural Interview were found.

Mechanisms to assist consistent administration of the interview and procedures have not been developed (G. van Ooyen, personal communication, 14 September, 2007).

## **Physical Competence Test (PCT)**

### **Meets standards**

Police Physical Education Officers (PEOs) are certified and reviewed by senior PEO instructors to ensure administration, scoring and interpretation consistently follows standardised procedures. In addition, the final PCT (completed by Recruits) is held at the Royal New Zealand Police College using standardised equipment and procedures and is conducted by a senior PEO (G. Wallwork, personal communication, 27 September, 2007).

Clear guidelines on how the confidentiality of assessment results is maintained and how the results will be used are provided to applicants and assessors.

Reassessment policies require the PEO conducting the assessment to determine an appropriate amount of time between a failed test and retesting (G. Wallwork, personal communication, 1 October, 2007). While no specific guidelines are documented, the policy is designed to enable a PEO to tailor the retesting time to best suit the individual's situation (e.g. level of fitness, gap between the result achieved and the pass mark, etc).

Clear and comprehensive administration instructions were available to assist assessors to conduct assessments, rate applicant performance and apply reassessment policies and procedures when required.

### **Scope for improvement**

Variation of testing requirements depending on facilities available. PCT guidelines provide the option of a long (outdoor) course and a short (indoor) course, yet no evidence of the impact of the different courses on assessment standards was provided (Physical Education and Defensive Tactics Section, 2004).

## **First Aid Certificate**

### **Meets standards**

Consistency of the national application with baseline standards is supported by the availability of specific documented standards which a First Aid Certificate must meet (New Zealand Police, 2001b).

## **Defensive Driving Certificate**

### **Meets standards**

Consistency of the national application with baseline standards is supported by the clear requirement that a Defensive Driving Certificate must meet Land Transport New Zealand standards.

## **Typing and Computer Skills Test**

### **Scope for improvement**

Limited documentation of administration procedures, resources required and the environment within which the assessment is to be administered is likely to reduce the consistency with which the Typing and Computer Skills Test is conducted, despite the observation that the test is straightforward and practical to administer and score.

## **SCOPE**

### **Meets standards**

Consistency of the national application with baseline standards is supported by the availability of clear instructions for administering the Pre-Course Reading Test.

### **Scope for improvement**

The necessary experience, qualification or training required of Police staff responsible for supervising and assessing SCOPE applicants was not described. Comments from District Recruitment Officers indicated that in their experience a variety of attitudes to SCOPE existed in different Stations, which in some instances had a negative impact on the motivation of the applicants involved.

Procedures to ensure the confidentiality of observations and ratings, and the security of interview materials, were not found.

Administration procedures provide insufficient guidance to workplace supervisors to ensure consistent administration of all aspects of the SCOPE assessment (e.g. re-sit policies for the Pre-Course Reading Test, the managing of non-completion of SCOPE hours, and Supervising Officer responsibilities).

A lack of clarity in the documentation provided regarding the role of SCOPE is likely to reduce the consistency with which SCOPE is used nationally. While it is described as a framework for logical investigation into what it means to be a Police Officer (G. van Ooyen, personal communication, 14 September, 2007), the applicant material says the “failure to complete the SCOPE programme will result in you not becoming eligible for selection” (New Zealand Police, 1997b, p. 2). The same documentation further reinforces SCOPE as an assessment tool by highlighting the need for applicants to pass the Pre-Course Reading Test (and providing an opportunity to re-sit if not passing first time).

No mechanism to ensure national consistency in how SCOPE is conducted was described in the documentation provided (e.g. monitoring and review procedures or a designated individual[s] with national co-ordination responsibilities).

## Reference Checks and Home Visit

### Meets standards

Administration instructions provide reasonable guidance regarding the range of information that checks and visits should collect.

### Scope for improvement

Policies for the training, accreditation, or required qualifications and experience of staff conducting checks and visits were not found.

Reference checking and home visiting practices across Districts appear to be variable, so reducing the consistency with which applicants are assessed. Discussions with staff indicated that a range of approaches is taken to how the checks and visit are conducted and when. For instance:

- › In one District, home visits are conducted only when specific concerns about an applicant have been identified. In another District, checks and visits are conducted for all applicants by full-time contracted personnel specialising in these tasks.
- › The reference checks and home visits conducted by some individuals just focus on the information provided by the 15FQ+, whereas others cover all the areas suggested in the Referee Interview Guide.
- › A range of perceptions appear to exist regarding the necessity for conducting international referee checks. In addition, methods for conducting these checks, and the agencies or other resources that can be used to assist, did not appear to be commonly understood.

No mechanism ensuring national consistency of the Reference Checks and Home Visit assessment was described in the documentation provided (e.g. monitoring and review procedures or a designated individual[s] with national co-ordination responsibilities).

Policies and guidelines to support consistent administration of the check and visits, and the security of referee checking and home visit guidelines, were not found (G. van Ooyen, personal communication, 18 September, 2007).

## Final Medical Clearance

### Meets standards

The policy requiring the use of Police certified, registered medical practitioners and Police certified examiners in specialities such as ophthalmology to conduct the Final Medical Clearance helps ensure consistency of standards (New Zealand Police, 2007b).

Guidelines to ensure the confidentiality of information collected, appropriate access to results and how the results would be used in the decision-making process were included in the documentation provided to applicants and certified examiners.

### Scope for improvement

While it is required that Final Medical Clearance assessments are conducted by Police certified examiners, this requirement may not be being met consistently. Discussions with District Recruitment Officers indicated that certified practitioners were not always available and in



some cases Final Medical Clearances are being conducted by uncertified General Practitioners (e.g. a recruit's own General Practitioner or a General Practitioner who is available at the time). While the specific accreditation procedure and standards were not reviewed, it would appear that using practitioners who are not certified by Police risks inconsistency in the application of assessment standards. District Recruitment Officers have in the past observed anomalies in the Clearances provided from practitioners not certified by Police.

A system to ensure national consistency of the Final Medical Clearance assessments was not evident in the documentation provided (e.g. a designated individual[s] with national co-ordination responsibilities, and documented monitoring and review procedures). While a medical practitioner (Dr Drummond) is contracted by Police to conduct training, and a senior Officer has responsibility for managing that relationship and assisting with assessment in the Districts, how this process worked and ensured consistency in standards of assessment was not described (M. Sutorius, personal communication, 26 September, 2007).

## Recruit training assessment

### Written Summative Assessments

The three Written Summative Assessments were considered together as the approach taken for each was the same.

#### Meets standards

Instructors are required to achieve competence consistent with NC5564, the National Certificate in Adult Education and Training (Level 4) (A. Richards, personal communication, 17 September, 2007).

Policies regarding the security and integrity of assessment materials are clear and specific.

The confidentiality of assessment results is guided by clear policies and regular quality assurance audits of assessment practices.

Reassessment policies are clear and based on sound rationale (Royal New Zealand Police College, 2004).

Sound administration procedures.

- > Examination administration procedures, including specific instructions for Recruits and resources required, are clearly documented and require the examination supervisor to 'sign off' completion of procedures.
- > Assessment policies require assessors to follow a quality control procedure for the marking, interpretation and moderation of assessment results where different stages of the procedure are reviewed and 'signed off' by senior staff and quality control specialists.
- > Policies describing procedures for academic dishonesty and communication of results are clearly documented.

### **Scope for improvement**

Recording practices resulting in errors in the recording of assessment scores (e.g. a number of graduating Recruits whose recorded scores did not meet the required pass marks).

### **Practical Summative Assessments**

The three Practical Summative Assessments were considered together as the approach taken for each was the same.

#### **Meets standards**

Instructors are required to achieve competence consistent with NC5564, the National Certificate in Adult Education and Training (Level 4) (A. Richards, personal communication, 17 September, 2007).

Policies regarding the security and integrity of assessment materials are clear and specific.

The confidentiality of assessment results is guided by clear policies and regular quality assurance audits of assessment practices.

Reassessment policies were clear and based on sound rationale (Royal New Zealand Police College, 2004).

Sound administration procedures.

- › Administration of practical assessments is clearly described and requires completion of standard checklists by the Officer in Charge of the assessment. Administration instructions include the briefing and debriefing of Recruits, and venue, resource and health and safety requirements. Rehearsals are conducted prior to the examination to ensure role-players and markers understand their roles and responsibilities.
- › Assessment policies require assessors to follow a quality control procedure for the marking, interpretation and moderation of assessment results where different stages of the procedure are reviewed and 'signed off' by senior staff and quality control specialists.
- › Policies describing procedures for communication of results and addressing academic dishonesty are clearly documented.

### **Scope for improvement**

Recording practices resulting in missing data and errors in the recording of assessment scores (e.g. a number of graduating Recruits whose recorded scores did not meet the required pass marks).

### **Driver Training**

#### **Meets standards**

Policies regarding the security of written assessment materials are clear and specific.

The confidentiality of assessment results and the storage of electronic and 'hard-copy' information on recruit assessments are guided by specific policies (Professional Police Driver Programme, 2007).

Reassessment policies were clear and based on sound rationale (i.e. designed to ensure mastery of essential elements of safe firearms management).

Sound administration procedures. Driving assessments are guided by checklists, specific performance standards and a description of the level of achievement required to pass the assessment. Administration instructions include assessment policies, procedures for recording and communicating results, procedures for dealing with variations in training and assessment processes, conditions under which the assessment process may vary (e.g. aegrotat assessment) and procedures for initiating moderation processes.

### **Scope for improvement**

While driving instructors are required to achieve competence consistent with NC5564, the National Certificate in Adult Education and Training (Level 4) and undergo a structured 'Driver Tutors Certification' (G. Northcott, personal communication, 5 October, 2007), RNZPC managers recognise the importance of developing and implementing a more rigorous specialist qualification, to ensure greater consistency in training and assessment of Recruits (S. A. Bruce, personal communication, 1 October, 2007).

## **Firearms Training**

### **Meets standards**

Assessors are certified New Zealand Police Firearms Instructors (M. Stonyer, personal communication, 28 September, 2007) and are required to achieve competence consistent with NC5564, the National Certificate in Adult Education and Training (Level 4).

The confidentiality of assessment results and the storage of electronic and 'hard-copy' information on recruit assessments are governed by RNZPC policies.

Reassessment policies were clear and based on sound rationale (e.g. Staff Safety Tactical Training, 2005a; Staff Safety Tactical Training, 2005c).

Sound administration procedures. Firearms assessments are guided by a detailed and extensive Facilitator's Guide describing specific performance standards, the levels of competence required to achieve a pass in the different elements of the assessment, and assessment procedures (Staff Safety Tactical Training, 2005c).

### **Scope for improvement**

Procedures to moderate assessments to help ensure consistency of marking standards were not observed in the documentation provided. For example, the Facilitator's Guide for the Bushmaster M4A3 rifle stated in respect of the written assessment, "In a dispute, the assessor's decision on the validity of answers is final" (Staff Safety Tactical Training, 2005c, p. 95).

## **Defensive Tactics**

### **Meets standards**

Assessors have completed a 5 week Staff Safety Tactical Training Instructors Course (Staff Safety Tactical Training, 2005d) and are required to achieve competence consistent with NC5564, the National Certificate in Adult Education and Training (Level 4).

The confidentiality of assessment results and the storage of electronic and 'hard-copy' information on recruit assessments are governed by RNZPC policies.

Policies regarding the security and integrity of assessment materials are clear and specific (Staff Safety Tactical Training, 2005b).

Reassessment policies were clear and based on sound rationale (Staff Safety Tactical Training, 2005b).

Sound administration procedures. Defensive tactics assessments are guided by a detailed Facilitator's Guide describing specific performance standards, the levels of competence required to achieve a pass in the different elements of the assessment, and assessment procedures (Staff Safety Tactical Training, 2005b).

## Workplace programme assessment

### Standards Assessment

The ten Standards assessments are considered together as the approach taken for each and method of assessment is the same.

#### Meets standards

Assessors are required to undertake RNZPC assessor training to achieve competence consistent with NZQA Unit Standard 4098 (A. Richards, personal communication, 4 October, 2007).

Consistent practice is also supported by the National Co-ordinator of the Probationary Constable Workplace Assessment Programme, who is responsible for ensuring national consistency (A. Richards, personal communication, 28 September, 2007).

Policies regarding the security and integrity of assessment materials are clear and specific (Recruit Training Group, 2006).

The confidentiality of assessment results is guided by RNZPC policies.

Reassessment policies were clear and based on sound rationale (Recruit Training Group, 2006).

Sound administration procedures.

- > Administration of Standards assessments is clearly described and utilises standard checklists to assist assessors prepare for an assessment, brief the Probationary Constable, evaluate and rate the evidence provided and/or performance observed, and provide feedback on the assessment results.
- > Assessment policies require assessors to follow standard practices regarding the marking and interpretation of assessment results.
- > A comprehensive moderation programme is used to help ensure consistency of marking standards. The programme includes regular scheduled moderation visits to workplace assessors by an independent National Moderator, an annual review of moderation standards and processes at a National Workplace Assessors Conference, and regular monitoring of the moderation programme by the National Co-ordinator of the

Probationary Constable Workplace Assessment Programme (Recruit Training Group, 2005b).

- › Policies describing procedures for communication of results and addressing academic dishonesty are clearly documented.

## **LEGL 114: Introduction to criminal law and problem solving**

### **Meets standards**

University staff responsible for the assessment are qualified and experienced lawyers.

The security and integrity of assessment materials is governed by Victoria University guidelines.

Confidentiality of assessment results is maintained by adherence to Victoria University's VPEP Quality Assurance Plan (J. Ramshaw, personal communication, 26 September, 2007).

Reassessment policies were clear and based on sound rationale (Victoria Police Education Programme, 2007).

Administration procedures were comprehensive and considered appropriate to the nature of the assessment involved.

## Summary evaluation of consistency of application

The level of consistency with the third baseline standard is summarised in the table below.

### Scale

<b>High =</b>	Close consistency between the assessment method and the baseline standard, indicating that a method is likely to be consistently applied by Police.
<b>Reasonable =</b>	Reasonable consistency between the assessment method and the baseline standard, indicating that the method is likely to be consistently applied, but that there are specific opportunities to enhance the consistency of application.
<b>Low =</b>	A low level of consistency between the assessment method and the baseline standard, indicating that Police cannot be confident that the method is being applied consistently.

<b>Recruitment Assessment Tools</b>	<b>High</b>	<b>Reasonable</b>	<b>Low</b>
Health Questionnaire	■		
Offences Clearance Check		■	
Swimming Certificate of Competency	■		
Cognitive Ability Testing (GRT2)		■	
Personality Testing (15FQ+)		■	
Physical Appraisal Test (PAT)			■
Behavioural Interview			■
Physical Competence Test (PCT)	■		
First Aid Certificate	■		
Defensive Driving Certificate	■		
Typing and Computer Skills Test			■
Reference Checks and Home Visit			■
SCOPE			■
Final Medical Clearance		■	
<b>Recruit Training Assessment Tools</b>	<b>High</b>	<b>Reasonable</b>	<b>Low</b>
Written Summative Assessments	■		
Practical Summative Assessments	■		
Driver Training		■	
Firearms Training	■		
Defensive Tactics	■		
<b>Workplace Programme Assessment Tools</b>	<b>High</b>	<b>Reasonable</b>	<b>Low</b>
PCWAP Standards	■		
LEGL 114	■		

An overall judgement of how well each assessment method was considered to meet the baseline standard was made based on the proportions of 'strong points' and 'scope for improvement' points, the impact these points were likely to have on the accuracy of assessment, and the nature of the instrument (e.g. the rigour expected of the administration, scoring and interpretation of a psychometric test was considered higher than that expected of a Swimming Certificate). These judgements are not definitive categorisations of the assessment tools, but judgements designed to assist the reader gain an overall picture of the conclusions.

As seen in the above table, the consistency with which the assessment tools were applied varied widely in the assessment of applicants but was consistently high for those used to assess Recruits during training and Probationary Constables during the Workplace Programme.

The key themes arising from the evaluation of how assessment tools were identified and applied are as follows:

- › Approximately one third of the assessment tools used to assess applicants (i.e. five out of fourteen) lacked sound mechanisms to ensure nationally consistent administration, scoring and interpretation of results, reducing the ability of these tools to provide comparable assessment across applicants. The most common fault was a lack of complete guidelines to ensure consistency of administration and interpretation (e.g. Behavioural Interview, Typing and Computer Skills Test, Reference Checks and Home Visit and SCOPE). Different equipment or venues had some impact on the Swimming Certificate and PAT.
- › Inconsistent administration practices were identified in the PAT, Reference Checks and Home Visit, and the SCOPE exercise, inconsistencies likely to be due to differences in District policies or the attitudes of individuals conducting the assessment.
- › The tools used to assess Recruits during training at RNZPC and those used to assess Probationary Constables completing the Workplace Assessment Programme largely matched the criteria under this standard. Characteristics of these tools were consistent administration practices, and sound practices to ensure security of testing materials and equipment and confidentiality of test results.
- › Errors in the recording of GRT2 assessment results were noted (Hattie, 2007) and in data-sets supplied describing Written Summative and Practical Summative Examination results. Hattie (2007) and Police staff assisting the author had difficulty matching different sets of assessment data, indicating the lack of a common, robust data-base of assessment results. The lack of an accurate central recording mechanism limits the ability of Police to accurately monitor assessment trends and to conduct empirical research to understand and improve the effectiveness of the assessment tools.
- › Issues specific to a single assessment tool included users of the 15FQ+ not all having the formal qualifications and experience typically required for users of this test, a situation which risks users not knowing when they need assistance and not knowing how to deal appropriately with the unexpected reactions of applicants.

## 9. References

- American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Anderson, T., & Penny, T. (2003). *An evaluation of the length of the 19 week Recruit Training Course at the Royal New Zealand Police College*. Wellington: NZ Police.
- Annan, W. (2006). *New Zealand Police: Recruitment standards review*. Wellington: NZ Police.
- Annan, W. (2007a). *Recruitment applicants: Policy on proven charges* (Police Executive Meeting paper 07/73). Wellington: NZ Police.
- Annan, W. (2007b). *Report to the Minister of Police: Police Recruits with previous offences*. Wellington: NZ Police.
- Australian Psychological Society. (2002). *Ethical guidelines (4th ed.)*. Melbourne: Australian Psychological Society.
- Barrett, P. (2005a). *Optimizing Police Officer selection*. Auckland: Barrett Consulting Services.
- Barrett, P. (2005b). *Predicting training success with GRT2 test records: Some initial investigations using 126 officer records*. Auckland: Barrett Consulting Services.
- Barrett, P. (2006). *The use of standardized-normatively scaled scores within a performance orientated selection process*. Auckland: Barrett Consulting Services.
- Black, J. (1995). *Recruit psychological testing project: Principles of psychological testing and summary of research findings*. Wellington: NZ Police.
- Carson, R. (2003). *Pursuits: The case for change*. Report to the Commissioner of Police. Wellington: Office of the Commissioner of Police.
- Chernyshenko, S. (2005). *Report on the psychometric evaluation of the General Reasoning Test (GRT2) for measurement equivalence across ethnic and gender groups*. Wellington: Drasgow Consulting Group.
- Defensive Tactics Section. (2003). *Police trainee required competencies*. Porirua: Royal New Zealand Police College.
- Duggan, M. (2000). *The validity and value of recruitment and selection: The relationship and the prediction of training and job performance* (Report prepared for NZ Police). Wellington: Author.
- Firearms Training Section. (2005). *Interim amended firearms instructors manual*. Porirua: Royal New Zealand Police College.



- Hattie, J. (2007). *Police standards and assessment practice (to Probationary Constable level)*. (Report prepared for NZ Police). Wellington: Author.
- Initial Training Group (2005a). *Generic Terms of Reference of the Curriculum Reference Group*. Porirua: Royal New Zealand Police College.
- Initial Training Group (2005b). *Defensive tactics instructor Level One: Course management package*. Porirua: Royal New Zealand Police College.
- Initial Training Group (2007a). *Recruit induction book*. Porirua: Royal New Zealand Police College.
- Initial Training Group (2007b). *Book One Part A: Information gathering and Police documentation*. Porirua: Royal New Zealand Police College.
- International Testing Commission (2000). *International Guidelines for Test Use*. Stockholm: Author.
- Kendall, I., Jenkinson, J., de Lemos, M., & Clancy, D. (1997). *Supplement to guidelines for the use of psychological tests*. Melbourne: Australian Psychological Society.
- Love, H. L., & Whittaker, W. (Eds). (1997). *Practice issues for clinical and applied psychologists in New Zealand*. Wellington: New Zealand Psychological Society.
- New Zealand Council for Educational Research (2006). *Human resources and psychological assessments: 2006 - 2007*. Wellington: Author.
- New Zealand Police (1998). *Project Baseline: A new Pay-fixing fixing Programme for New Zealand Police*. Wellington: Author
- New Zealand Police. (1997a). *Recruiting enquiries: Home visits*. Wellington: Author.
- New Zealand Police. (1997b). *SCOPE*. Wellington: Author.
- New Zealand Police. (2001a). *New Zealand Police position description: Traffic Constable*. Wellington: Author.
- New Zealand Police. (2001b). *Seeking a career in the New Zealand Police*. Wellington: Author.
- New Zealand Police. (2004). *Request for proposal (NZ Police as purchaser) RFP TN/04/16 (Contract)*. Wellington: Author.
- New Zealand Police. (2005a). *Application to join the New Zealand Police: POL21*. Wellington: Author.
- New Zealand Police. (2005b). *Formal employment interview*. Wellington: Author.
- New Zealand Police. (2006a). *Health questionnaire*. Wellington: Author.
- New Zealand Police. (2006b). *The guidelines for GRT2 tests* (internal procedures description). Wellington: Author.
- New Zealand Police. (2007a). *Recruit medical screening procedures*. Wellington: Author.
- New Zealand Police. (2007b). *Ophthalmic standards for New Zealand Police: Accredited vision examiner protocol*. Wellington: Author.

- New Zealand Police. (2007c). *Auckland Police Recruitment: Operations Manual*. Auckland: Auckland Police Recruitment.
- New Zealand Police. (2007d). *Guidelines for practitioners: Sworn applicant process - vetting and validation*. Wellington: NZ Police.
- O'Callaghan, T. (2003). *NZ Police: Assessment testing for recruitment and selection project draft report*. Wellington: OPRA Consulting Group.
- OPRA Consulting (2004a). *Detailed project plan: Introduction of GeneSys assessment tests for Police Recruits*. Wellington: Author.
- OPRA Consulting (2004b). *Development of a 15FQ+/GRT2 ideal role profile and validation study*. Wellington: Author.
- OPRA Consulting (2004c). *15FQ+ interview questions*. Wellington: Author.
- Physical Education and Defensive Tactics Section. (2004). *Physical Competence Test*. Porirua: Royal New Zealand Police College.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory*. Mahwah, NJ: Erlbaum.
- Professional Police Driver Programme. (PPDP). (2005). *Workplace assessment record book*. Porirua: Royal New Zealand Police College.
- Professional Police Driver Programme. (PPDP). (2007). *Initial driver assessment record book*. Porirua: Royal New Zealand Police College.
- Psychometrics Limited (2002). *The 15FQ+ Technical manual*. United Kingdom: Author.
- Psytech International Limited. (2007). *General, Graduate & Internet Reasoning tests'': Technical Manual*. Pulloxhill, United Kingdom: Author.
- Recruit Training Group. (2005a). *Probationary Constable Workplace Assessment Programme: Assessors' handbook*. Porirua: Royal New Zealand Police College.
- Recruit Training Group. (2005b). *Probationary Constable Workplace Assessment Programme: Moderation plan*. Porirua: Royal New Zealand Police College.
- Recruit Training Group. (2006). *Workplace assessment record book: Probationary Constable Workplace Assessment Programme*. Porirua: Royal New Zealand Police College.
- Recruit Training Group. (2007). *Probationary Constable Workplace Assessment Programme: Assessor's assessment questions*. Porirua: Royal New Zealand Police College.
- Research and Development Group. (1999). *Police recruit training programme: pre-course study*. Training Service Centre, Porirua: Royal New Zealand Police College.
- Robertson, I. T., & Smith, M. (2001). *Personnel selection*. Journal of Occupational and Organizational Psychology, 74, 441-472.
- Robinson, P. (2006). *Review: Police recruit medical screening standards*. Wellington: NZ Police.

- Royal New Zealand Police College. (2002). *Performance management guide for Probationary Constables*. Porirua: Author.
- Royal New Zealand Police College. (2003). *Interim report: Professional development of instructional staff in initial recruit training: Research and development project RD17*. Porirua: Author.
- Royal New Zealand Police College. (2004). *Overview of recruit training*. Porirua: Author.
- Royal New Zealand Police College. (2007). *Recruit induction book*. Porirua: Author.
- Schippmann, J. S., Ash, A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., Kehoe, J., Pearlman, K., Prien, E. P., & Sanchez, J. (2000). The practice of competency modelling. *Personnel Psychology*, 53, 703 - 740.
- Society for Industrial & Organisational Psychology. (SIOP). (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green, USA: Author.
- Staff Safety Tactical Training. (2003). *Police Trainee Required Competencies*. Porirua: Royal New Zealand Police College.
- Staff Safety Tactical Training. (2005a). *Bushmaster M4A3: Operator's handbook*. Porirua: Royal New Zealand Police College.
- Staff Safety Tactical Training. (2005b). *Defensive tactics: Facilitator's guide*. Porirua: Royal New Zealand Police College.
- Staff Safety Tactical Training. (2005c). *Bushmaster M4A3 training: Facilitator's guide*. Porirua: Royal New Zealand Police College.
- Staff Safety Tactical Training. (2005d). *Defensive tactics instructor level one: Course management package*. Porirua: Royal New Zealand Police College.
- State Services Commission. (1998). *Managing to be fair: Human rights and Public service Employment*. Wellington, New Zealand: Author.
- Training Service Centre. (2001). *Initial Police training curriculum: Years 1 - 3*. Porirua: Royal New Zealand Police College.
- Trappitt, D. E (2007). *New Zealand Police Training 2007: A study of organisational preparedness of Police Training in support of the Police Strategic Plan*. Wellington: New Zealand Police.
- Tyler, G. (2003). A review of the 15FQ+ personality questionnaire. *Selection and Development Review*, 19, 7 - 11.
- University Teaching Development Centre. (2004). *Improving Teaching and Learning: The Mechanics of Assessment*. Wellington: Victoria University of Wellington
- Victoria Police Education Programme (VPEP). (2007). *LEGL 114 Introduction to criminal law and problem solving: Course Information*. Wellington: Faculty of Law, Victoria University.

Wilson, N. C., & Toomey, M. (1990). *Final report on a review of the physical component of the required fitness level for Police Recruits*. Dunedin: Human Performance Centre, Otago University.

Wilson, N. C., & Toomey, M. (1993). *Review of the entry standards for New Zealand Police Recruits*. Dunedin: Human Performance Centre, Otago University.

## 10. Appendix

This Appendix contains the independent advice on the reports prepared by Senior Sergeant Iain Saunders and Wayne Annan as input to the Advice on the processes New Zealand Police use to determine minimum standards, determine assessment methods, and to apply these methods consistently

## POLICE STANDARDS AND ASSESSMENT PRACTICE (TO PROBATIONARY CONSTABLE LEVEL)

*Independent advice on the reports prepared by Senior Sergeant Iain Saunders & Wayne Annan as input to the work of Dr George on the processes NZ Police use to determine minimum standards, determine assessment methods, and to apply these methods consistently<sup>1</sup>*

John Hattie  
Professor of Education, University of Auckland  
August, 2007

The Minister of Police has requested independent expert advice on the standards and assessment practices for police applicants, police recruits and probationary constables to determine whether they meet best practice. “Terms of Reference” (TOR) for a more complete overview have been agreed. This paper has a more narrow focus – it provides an independent study of the efficacy of the data, methodology and conclusions offered in the reports written by Senior Sergeant Iain Saunders, RNZPC (“Recruit training and the strategic direction of Policy – Standards and performance”), and Wayne Annan, General Manager Human Resources (“Analysis of staff with low stanine score entering RNZPC”). A one-page summary of Key Points is attached at the end of this Report.

A first draft of this report was provided to Annan and Saunders, and I met with them to seek clarification of some issues, invite comments, and I thank both for their involvement and assistance. Subsequent to completing a first draft of this report, a new 44 page untitled paper was provided by Wayne Annan, and several emails provided to me by Iain Saunders. I also learnt that both the above papers were internal documents in very draft stages and never meant to be the basis of a major debate. Saunders paper, for example, was a first draft sent to his superior for comment as to whether he was moving in the right direction but soon was escalated out of his control – he indicated to me that he would have wished to present the paper in a more refined and final draft form, cross checking evidence, seeking others input etc., – this draft nature is evident in the paper and thus needs to be taken into consideration when reviewing and acting with this paper. Annan also expressed concern about his initial paper, which he stated was written prior to his seeing the Saunders report – and was developed in reaction to a brief comment in the PoliceNews.

It is critical to note, that while an independent review often aims to resolve differences, it is more important to move forward; hence this review is aimed, via its recommendations, to feed into the TOR about processes Police use to determine minimum standards, determine assessment methods, and to apply these methods consistently across New Zealand.

### ***There Are Seven Matters I Wish to Discuss in Relation to the Two Papers Provided:***

- A. *The lack of any reference to a job analysis;*
- B. *Is general mental ability a useful predictor in police work?*
- C. *The choice of assessment tools described in these papers;*
- D. *The misuse of stanines;*
- E. *The lack of discussion about standard setting for setting cut-scores;*
- F. *The question of whether standards of police recruits have declined?*
- G. *The process for conducting research within the Police.*

1. I thank Jeanette Schollum, Wayne Annan, and Iain Saunders for their openness, input, critique and assistance. I, however, take full responsibility for this report.

### A. The Lack of Any Reference to a Job Analysis

One of the difficulties in resolving the issues highlighted in the papers by Annan and Saunders, and in general relating to police recruitment methods, is that there is no reference to a job analysis as a basis for the selection process for police recruits (Brannick & Levine, 2002; McEntire, Dailey, Osburn, Mumford, 2006; Wilson, 2007). The results of a job analysis (particularly of the tasks encountered in the first years of policing) would permit the Police to specify the qualifications and the desired level of attributes needed by police recruits – and provide the critical validity criteria to evaluate any selection tools and methods. I have been told that there is an understanding that such a job analysis is considered important, that there are job analyses for specific positions in the police, but none for the 1<sup>st</sup> years in the profession.

Completing a job analysis prior to developing criteria or a selection mechanism for selecting recruits is consistent with sound professional practice. In the USA, as early as 1985 the Standards for Educational and Psychological Testing produced by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education (AERA, APA, NCME, 1985), and the 1978 *Uniform Guidelines on Employee Selection Procedures* (U.S. Equal Employment Opportunity Commission, EEOC, 1978), which have been endorsed and promulgated by four federal government agencies—the U. S. Equal Employment Opportunity Commission, the U. S. Department of Labor, the U. S. Civil Service Commission, and the U. S. Department of Justice—defined principles and procedures that are applicable to selection methods and devices used by public agencies and professional organizations for licensure or certification of those who wish to practice in an occupation or a profession. The *Guidelines* apply to tests and other selection procedures that are used as a basis for any employment decision. These *Guidelines* have been updated and refined but the fundamental principles have remained throughout many professions.

The *Guidelines* identify job analysis as the *sine qua non* of procedures for amassing content-related validity evidence for recruiting and licensure procedures. A job analysis must include “an analysis of the important work behaviour(s) required for successful performance and their relative importance” (Section 14. C. 2.). The *Guidelines* evidence a substantial preference for assessments that incorporate actual work behaviours or work samples, although they permit use of recruitment tests of knowledge, skills, and abilities that are prerequisite to successful job performance. According to Section 14. C. 4.: “As the content of the selection procedure less resembles a work behaviour, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.”

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985) include various standards that address validation of certification and licensure procedures and these are particularly pertinent to the job relevance criteria defined in the *Guidelines*.

*Standard 11.1 The content domain to be covered by a licensure or certification test should be defined clearly and explained in terms of the importance of the content for competent performance in an occupation. A rationale should be provided to support a claim that the knowledge or skills being assessed are required for competent performance in an occupation and are consistent with the purpose for which the licensing or certification program was instituted. (Primary)*

*Standard 11.2 Any construct interpretations of tests used for licensure and certification should be made explicit, and the evidence and logical analyses supporting these interpretations should be reported. (Primary)*

Job analyses provide the primary basis for defining the content domain. The claim that a particular skill is necessary for competent practice in the police profession involves inferences that should be supported by evidence and logical analysis. Good performance on a certification examination should not require more reading or writing ability, for example, than

is necessary in the occupation. The job analysis procedures used in establishing the content-related validity of a test can also contribute to the construct interpretation. One may show, for example, that qualified experts helped to define the job, identify the knowledge and skills required for competent performance, and determined the appropriate level of complexity at which this knowledge and skills should be assessed.

When content-related evidence is used as the principal means of supporting a claim to the validity of a certification or licensure procedure that assesses knowledge, the cited standards and guidelines suggest that validation consists of amassing support for three related claims: (1) the knowledge required by the recruitment procedure must be required for safe and effective practice in the occupation or profession for which certification or licensure is sought; i.e., a practitioner who did not possess that knowledge would be dangerous to the public; (2) the items or exercises that compose the assessment procedure must assess the required knowledge, and nothing else; and (3) persons who provide judgements concerning the first two claims are qualified to do so.

More recently, the US based database O\*Net has been developed to provide a standardized, comprehensive and online system for assisting in job analysis (<http://online.onetcenter.org/link/summary/33-3051.01>). It could be used as a starting point to describe most any job in the organizations of today and in the future (Mariani, 2001) – including New Zealand Police Officers. It outlines the major tasks of most professions, describes jobs in terms of the abilities necessary to do the work, provides a common language to describe and compare jobs, has determined a hierarchical structure of major aspects of the job and subsidiary skills to undertake these jobs, and links to metrics and methods to assist in furthering and refining the analysis of jobs. For example, for police recruits it specifies various tasks:

- Provide for public safety by maintaining order, responding to emergencies, protecting people and property, enforcing motor vehicle and criminal laws, and promoting good community relations.
- Identify, pursue, and arrest suspects and perpetrators of criminal acts.
- Record facts to prepare reports that document incidents and activities.
- Review facts of incidents to determine if criminal act or statute violations were involved.
- Render aid to accident victims and other persons requiring first aid for physical injuries.
- Testify in court to present evidence or act as witness in traffic and criminal cases.
- Evaluate complaint and emergency-request information to determine response requirements.
- Patrol specific area on foot, horseback, or motorized conveyance, responding promptly to calls for assistance.
- Monitor, note, report, and investigate suspicious persons and situations, safety hazards, and unusual or illegal activity in patrol area.
- Investigate traffic accidents and other accidents to determine causes and to determine if a crime has been committed.

And then it specifies the knowledge, skills, abilities, work activities, work contexts, job zones, job interests, work styles, work values, and wage and employment trends (see Appendix 1; although it is noted that there would need to be modifications to fit the NZ context and laws). This resource could be a valuable source to commence a job analysis of New Zealand Police (if there is none currently available) and would then allow a closer inspection of the match and predictability of any recruitment assessment systems and tools.

***Recommendation 1: It is recommended that the Police undertake a job analysis, with particular reference to the first years of policing, to provide the basis for then choosing selection tests and desired cut-scores on those tests to enter the profession.***



## B. Is General Mental Ability A Useful Predictor In Police Work?

The answer clearly depends on the outcomes of the above recommended job analysis. The Schmidt and Hunter research cited in the Saunders paper is among the best available on the issues of the place of mental abilities in employment and the evidence is convincing that general intellectual ability (in this case measured by reasoning tests) underlies success in many professions. The issue is not whether mental ability or reasoning is a useful predictor or critical aspect of ongoing police work (certainly it appears to be, see Appendix 1), but the issue is the *priority* of these proficiencies and the *minimum level* needed to undertake the work tasks in policing. I understand from other papers provided that this test is *not the only* screening tool; as should be the case. That the police are still using general ability measures as part of a profile should provide much confidence about the quality of recruits, provided the cut-score for these tests is at the defensible level of desired ability to undertake police work (which is addressed below).

There is the message in the Saunders paper that the police are “recruiting at around the 23<sup>rd</sup> percentile from the general population as a minimum standard” and thus “77 of every 100 members of the population they (encounter) would be better at reasoning tasks ... than the officer” (Saunders, p. 10). If reasoning was the “only” criterion for entry this may indeed be the case. But it is *not* the case that reasoning is the only criteria and a robust and dependable *profile* of the recruits should be (and *is*) the key set of attributes used for recruiting – realising that inclusion of cognitive attributes should remain a critical basis of the initial selection. An IQ at the 23<sup>rd</sup> percentile is about 90 – well within what would be considered “normal IQ” for the population. Imagine a recruit with IQ = 90, but superior personality and physical skills, compared to a recruit with IQ = 120, who is not very conscientious and barely passes the physical appraisal tests. Which is preferred? The key is the consideration across a *profile* of attributes, and the claim above implies that only one attribute is considered. (Further, the use of stanines means that a recruit in “Stanine 4” could score between 90 and 96, so there is no reason to immediately believe that the IQ is indeed 90 – it could be 96!)

There are many content validation studies of reasoning tests which show that they should not be confused with academic, numerical, vocabulary abilities – for example the most well known measure of reasoning, the Raven’s Progressive Matrices, require no knowledge of numerical, verbal or academic skills. Measures of fluid intelligence (such as those measured by the currently used reasoning tests) are far less reliant on school and background influences than crystallised intelligence measures.

There are probably many aspects of the current police job that would require assessments more directly tied to schooling and prior experience – such as writing (which seems missing in the selection profile), ability to read to a reasonable level (e.g., legal documents), and self-regulation (e.g., study skills). The Saunders report refers to literacy skills – and these, of course, are not measured by the reasoning tests; and if a job analysis indicates that literacy skills are needed (and specifies the level of skill required), then a literacy test could be a defensible addition to the selection tests.

The regression analyses in the Saunders report should provide some confidence in the use of these reasoning skills as powerful predictors of the “overall wing marks” or any final aggregate of assessments that leads to graduation from the training course. I would have preferred, however, to have seen all measures in the regression equation to show more relative importance of all selection criteria. If the Verbal reasoning measure remains the most powerful, then there could be a case for only using this measure and not using Numerical or Abstract reasoning, except perhaps as additional sources of information if Verbal reasoning is just below the cut-score – or they could all be retained if the job analysis indicates that all types of reasoning are key attributes of success in policing.

### C. The Choice of Assessment Tools Described in These Papers

The test in question is a general test of reasoning (Abstract, Numerical, Verbal). It appears that the General Reasoning Test Version 2 (GRT2) was introduced following a review – whereas previously the Australian Council for Education Research's Higher and Advanced Tests - Form ML/MQ (ACER ML/MQ) was used (O'Callaghan, 2003). The argument against continuing to use the ACER test was that it was more a measure of crystallised intelligence (but it has the same dimensions as the GRT2 test!); it has no New Zealand norms (but neither does the GRT2), and it is not computer administered. I understand that the ACER tests were exposed and this is a critical reason for changing – although the GRT2 in time will probably also be exposed. An item bank of reasoning items, administered via computer on a stratified random basis, would assist to avoid this exposure.

The GRT2 manual specifies norms based on the “general population”, telesales applicants, college students, customer service clerks, technical staff, financial consultants, HR professionals, and service engineers. The tests favour males on Numerical, and Abstract reasoning, but not on Verbal reasoning. There are very high correlations between subtests (Verbal – Numerical  $r = .60$ , Verbal – Abstract  $r = .56$ , Numerical - Abstract  $r = .65$ ).

The Chernyshenko (2005) report on the GRT2 based on 255 New Zealand Police recruits is exemplary. It clearly outlines the major considerations of the psychometric properties of the GRT2 and should provide much confidence in the dependability of these measures. The estimates of reliability are sufficiently high, the standard errors are appropriately small (about 2 on each test, such that a difference of  $2 \times 2 = 4$  score points can be used to discriminate), there is evidence of a single factor such that a total score on each test is interpretable, and there is no evidence of item bias with respect to gender or ethnicity. I would concur with the conclusion that the “GRT2 was judged to have adequate psychometric properties for applicant screening and selection purposes” (p. 32).

Compared to the norm sample in the GRT2 manual, New Zealand Police Recruits are similar to the population Numerical ( $d = -.17$ ), lower in Verbal ( $d = -.29$ ) and higher in Abstract reasoning ( $d = .33$ ).

	Manual Mean	NZ Police Recruits Mean	sd	Effect-size
Verbal	23.30	21.91	4.88	-0.29
Numerical	16.64	15.82	4.92	-0.17
Abstract	17.13	18.42	3.94	0.33

According to the document entitled the “New Zealand Police Recruitment Process” and “Rules for GRT2 tests”, the score on Abstract reasoning is used as a first criterion for selection, and if the score is less than Stanine 3, then the candidate proceeds no further. If in Stanine 3 the applicant is declined but can “come back when better prepared” – although another document provided called “The guidelines for GRT2, March 2006” contradicts this and allows those in Stanine 1 to 3 to be stood down but can “study and come back when better prepared”. Both documents state that and those with Stanine 4 “maybe – other variables to be considered, refer to Megan J for decision”. Such consideration includes scores on the other two reasoning tests, personality test scores, proven levels of motivation, determination, current job and exposure to report writing, working with numbers, etc. education/study history, interviews, PAT/PCT references, SCOPE etc. and on the whole a good candidate.

***Recommendation 2: It is recommended that there needs to be clear advice on the consequences of scoring in Stanine 1-3 (there needs to be resolution between the various “Rules” and “Guidelines” for minimal entry using the GRT2 tests).***

Then the Numeracy and Verbal scores (incorrectly called literacy in the document “New Zealand Police Recruitment Process”) are checked and the same criteria are set – candidates with scores below Stanine 3 are rejected in one document, but are stood down and can come back when better prepared. The regression in the Saunders paper may cause the selection panel to consider Verbal rather than Abstract reasoning first, but again a job analysis also should inform this decision.

#### **D. The Misuse of Stanines**

Stanines are normalized scale scores that have a mean of approximately 5 and a standard deviation of approximately 2 for a particular population of examinees; the resulting scores are then rounded to integers. Stanines are so often confusing to users that I always recommend NOT using them. Despite their seeming simplicity, the current two reports under consideration provide examples of how simplicity can lead to problems and misrepresentations.

##### *Stanines are not scores but bands of scores.*

Take, for example, Stanine 4 on the IQ scale. A police recruit in Stanine 4 could score between 90 and 96 – a large range. And it should not be assumed that because they are classified as falling in Stanine 4 the recruits are therefore at 90 or 96. Further, the difference *within* Stanine 4 is 6 IQ points, whereas the difference *between* Stanines 3 and 4 is 1 IQ point (i.e., 89 or 90). It does not seem reasonable to treat scores of 90 to 96 as indistinguishable, while at the same time treating scores that could differ by only one point as distinct – particularly when the standard error of measurement (in this case of 4) means that a score of 89 and 90 are not meaningfully different!

##### *The reference group is critical*

As Saunders notes, stanines allow a comparison to a reference group, and thus as the reference group changes so too does the meaning of the stanines; and as the representativeness of the group changes, so too does the meaning change. Barrett (2006) submitted a paper about the “Risks of using norms in the manner Police currently appear to.” He claimed that there “is clearly no justification whatsoever for using transformed scaled scores such as stens, T-scores, stanines, etc., in a performance-oriented selection process, except where the norms are properly representative, substantive in constituent number, and remain static” (p. 2). He argued against using population norms to form stanines (except as a “last resort”), he provided a compelling simulation showing how the same raw score can lead to different stanine/sten scores with major differences in interpretation (given that the raw score is no different this is not defensible). He showed that as extra candidates are added to the local norm group this can lead to wrong decisions to admit police recruits, and he concluded by claiming that the Police should work with non-banded scores (see also Schmidt, 1991; Schmidt & Hunter, 1995).

##### *The use of labels is arbitrary*

The use of words like “low average” for Stanine 4 is misleading. It means an applicant is scoring below 40% of the population of interest— if the reference group is Einstein-like physicists, I would be pleased to be at Stanine 4, but if the reference group is kindergarten graduates I would not be pleased— the interpretation is dependent on the reference group. In the jargon of standard setting, the aim is to set the cut-score for the “minimally competent” recruit and this requires a professional judgement. References to norm-distributed notions are misplaced in such selection decisions.

It appears that the database used for computing the stanines is built by OPRA every few years. It would be useful for Police Management to receive a report from this consulting company about any changes to the stanine cut-points, the implications of these changes, and any other psychometric information about the tests, selection cut-scores, and other psychometric analyses.

The Saunders report refers to the norm group for comparison appearing “to have been subject to regular fluctuations” (p. 13). He refers to Barrett’s paper which includes a simulation, not a statement about current policy practice in the reference group. He comments that the database includes updates of recruit scores – and a raw score of 18 falls in Stanine 3 in 2004, but currently it is in Stanine 4 (as would be predicted by the Barrett paper). This highlights the problems of using stanines and a shifting population/normative group. But, if the norm group is carefully built every few (and not every) years, is sufficiently large and robust, then these fears are misplaced.

*Stanines cannot be compared across different tests.*

The meaning of Stanine = 4 on one test can be quite different to Stanine = 4 on another test. The equivalence is only that they represent cuts in a distribution of scores within each test. A score at stanine 6 in Mathematics and at 8 in Reading are not directly comparable. If the comparison group in Mathematics is very competent in Mathematics, but not in Reading, then a Stanine of 6 in Mathematics is most impressive and probably indicates that the person’s competence in Mathematics is far superior relative to their competence in Reading. Similarly, if the comparison group is Reading is not very competent in Reading then a Stanine of 8 may not signify a high level of Reading.

*Differences in stanines are often difficult to understand*

The typical argument is that it is only when a person advances by 2 stanines can we be sure that there is growth. This can lead to absurd implications. For example, the difference between IQ 97 and 110 is *not* different (13 IQ points) but IQ 76 and 83 (7 IQ points) *is* statistically significantly different! A difference in IQ from 96 to 97 is a whole stanine different, but 97 to 98 is trivially different!

*Stanines increase errors of measurement*

One implication of using stanines is that the error of measurement can increase. For example, if the estimate of reliability for the raw scores is .90, then rounding to form stanines increased the mean-squared measurement error by about 42%—an unnecessary increase in error/reduction in measurement accuracy due to aiming for simplicity (see Kolen, 1988).

The message must be that the ***use of stanines should be stopped***. They lead to errors (measurement, typing/transposing, database errors), and a change in the reference group can have a profound on the decision to select or reject a candidate. Instead:

***Recommendation 3: It is recommended that the raw scores or an IRT-based proficiency estimate should be used—a cut-score established, and the relevant error of measurement (\*1.96,  $\alpha = .05$ ) used to make decisions about selection or rejection.***

### **E. The Lack of Discussion about Standard Setting for Setting Cut-Scores**

It seems that an arbitrary cut-score between the 3rd and 4th stanine has been chosen for entry. Notwithstanding the problems raised above, it is an indefensible use of stanines to establish the cut-score for the minimally competent recruit. ***Standard setting*** methods include now well-known processes for making decisions about where a cut-score should be placed to discriminate between those who are optimal to enter or those who should not enter police training. It seems that these methodologies have not been used to set cut-scores on the reasoning tests.

Setting performance standards for selection into police training is a process of eliciting reasoned judgments from experts who are (a) knowledgeable about the demands of the test or assessment for which a standard is to be set, (b) understand the meaning of scores at various levels on the scales used to summarize examinees' performances, and (c) fully comprehend the definitions of achievement associated with the performance standards that they have been asked to establish. As noted in the Barrett (2006) report it needs to be accompanied by appropriate psychometric data about actual performance.

It is important that the standard-setting method reflects the nature of the decision process, that it be replicable, and that there is evidence to support the intended interpretations and/or to refute competing interpretations (Kane, 1992; Shepard, 1993). It is not defensible to set up "committees" to debate issues, decide on standards and then get some buy-in from other groups. Such a method has no psychometric rigour, and is often swayed by the beliefs of a very small number of persons in the committee. Instead, standard-setting has become a major focus of many research studies, the basis of many court decisions, and there is a large body of literature on how to set standards (Cizek, 2001; Jaeger, 1989) and in recent years there have been comparisons between the standards set by different methods (Hattie & Brown, 2003; Plake, 1995).

Typically in standard setting methods, one or more panels of judges are assembled for the purpose of recommending what examinees should know and be able to do to achieve some valued end, and to also specify an associated score on a test that is regarded as an indicator of that requisite knowledge and ability. This associated score is commonly called a cut-score, but it is important to distinguish between the notion of cut-score (the score on, for example, some reasoning tests chosen to select or classify examinees with respect to the performance standard), and a performance standard ("the minimally adequate level of performance for some purpose," Kane, 1994, p. 425). Thus, cut-scores are points on the reasoning test scale, for example, that form boundaries between levels of performance, and performance standards are the specifications of the knowledge, skills, and abilities needed to accomplish various levels of performance.

There are many major methods of setting standards, such as the test-centered (judgements about items), examinee-centred (judgements about examinees), processing-centred methods (judgements about processing items), and the bookmark method. This is not the place to review these methods, but use of these methods, and a more detailed discussion can be found in Hattie and Brown (2003). In our own work in the development of the asTTle application, we have found much support for the Bookmark method (Lewis, Green, Mitzel, Baum, & Patz, 2003; Mitzel, Lewis, Patz, & Green, 2001) as it is intuitively easier to involve non-measurement specialists in this method, it leads to greater confidence in the agreed cut-score, uses the information from the psychometric attributes of the test to assist in setting the cut-score, and has much credibility in the measurement community.

To provide confidence in the selection cutting point, provide more rigour to keep constant the cut-score for selection across years, and to allow more dependable information about changing standards (or not) in the applicant pool:

***Recommendation 4: It is recommended that the Police use a standard-setting process (such as the Bookmark method) to determine the appropriate cut-score on the various assessments it is using in the selection of Police recruits.***

## F. Have Standards of Police Recruits Declined?

The major debate in these two papers relates to whether there is, or is not, a reduction in the quality of recent recruits during training. Evidence given for this *decline* include:

1. Saunders claimed that there was a 150% increase in demand for tutorial services over the past three years; an “increase from 7.5% to 13.5% in percentage academic fails on Exam one”/ “entering remedial training” for pre and post GRT2 introduction; and that there is a marked decrease in mean scores on the first written examination. This conclusion is based on the notion that there was a transition from the ACER to the GRT2 between Wings 219 to 222; and the average before Wing 218 was 83% and since Wing 223 was 75%. In Annan’s rebuttal paper, a table is provided showing that the GRT2 was first used with Wing 227 (N=3) and only from Wing 229 onwards was GRT2 the more common assessment than ACER. If Wing 229 is used as the transition then the averages change to 79% pre-227 and 79% (from ACER) post-227 – thus no differences. Further Annan claims that Saunders appears to use all recruits not just those that have graduated (and Saunders confirmed this was the case as his interest was in all recruits as some did not graduate because they were poorly prepared, failed aspects of training, resigned due to exam failure or pressure of training, etc.).
2. In rebuttal (via discussion and subsequent emails), Saunders indicated that the Wing number for dividing the testing methods was deduced from the database provided by PNHQ. Further, Saunders claimed that he was involved in a case of a recruit entered using ACER as early as Wing 223 (and ACER certainly was used by Auckland recruitment at that time, he claimed; Annan notes Wings 221 and 222 were tested with the GRT2 as part of pre-implementation, probably to evaluate the new test); that his (Saunders) choice related to when “GRT2 and associated standards” were introduced; that up to Wing 218 the standard was Stanine 5 not 4; and that he used the “scores used to select the member as reported to us on summaries” (see comment on these summaries below). The data provided to me by Michael Sutorius supports the conclusion that the GRT2 was “officially” introduced with Wing 227 (although it was used on a trial basis prior to this time).

There appears to be no disagreement that there has been a decline in the pass average for Examination One since Wing 215. The data in the two papers have close correspondence – the key issue is more *when* a trend downwards occurred. Annan agrees there is a trend downward (the “trend downward for Summative 1 Exam began at Wing 218”), and Saunders argues for a decline from some earlier group of Wings compared with a latter set of Wings. The regression coefficient across all Wings from the Saunders Graph is -.50 and from Annan is -.59 – the negative sign indicating evidence of a decline (and both are statistically significant:  $t=-2.78$ ,  $p = .011$ ;  $t = -3.61$ ,  $p<.001$ , respectively). Thus there is evidence in both papers that there is a decline – the point of contention is whether there is an “interruption” to the time series that can then be attributed to an event – such as the movement from ACER to GRT2, and/or a specific lowering of entry standards. *Overall, the changes in pass marks in Examination support the claim that there is a decline in performance and it could be attributed to the changing standards of the cohort* (or it may be attributed to the changing difficulty of the written examination although Saunders claimed that this written examination has remained relatively stable over time). Annan in his rebuttal claims that these changes may be related to more movement among the top students, a change in the number of questions, and/or a change in relative weightings. What is in dispute is not whether there was a decline in the performance in this first test, but the possible causes.

In the Annan rebuttal, there is a graph of “all assessments” from those entered with ACER and with GRT2 – and the final average percentage was 82% for those entering

under ACER and 83% for those entering under GRT2. This is evidence that the final decision to graduate a police recruit does not seem to differ as a function of any change in assessment (and any accompanying change in standards) – but this could be moderated by many factors besides entry assessment, such as the (enhanced) quality of the training course, and the effectiveness of the remedial assistance.

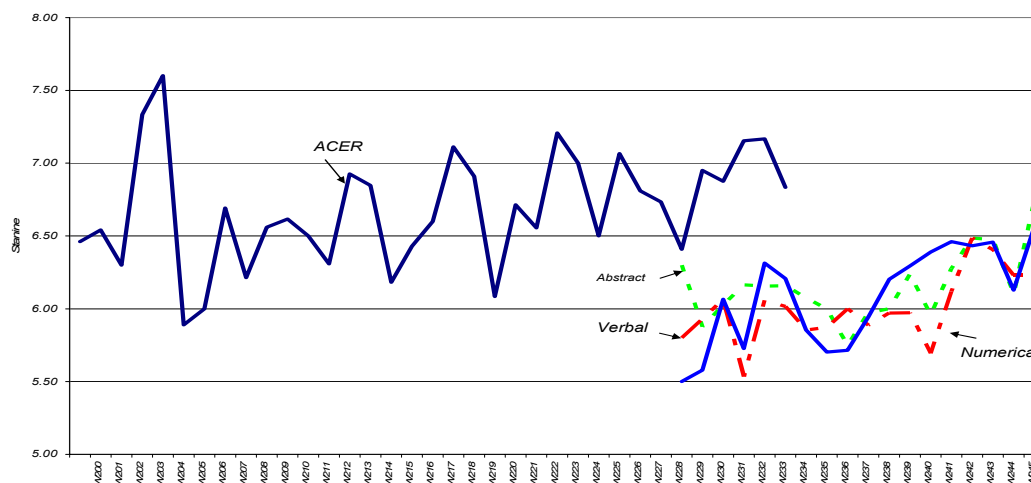
*There should be much comfort for the Police, however, that despite any changes in entry assessment and standards, the performance at the end of the training does not seem to relate to these changes.*

This evidence does indicate that statements about the “quality” of Police after success in the training program (as opposed to entry into this Police initial training course) should not be based solely on standards for initial selection *into* the profession, but should also take into account the performance in the exit examinations of the training program. It is certainly a widely debated notion as to whether entry or exit standards of training are more critical. My personal view is that a profession needs reasonable and appropriate (referenced to a job analysis) entry standards but the emphasis needs to be placed more on ensuring that there are appropriate exit assessments from initial training (referenced to clear Standards of entry into the next level - Constabulary workforce) that are enforced (i.e., those not passing are exited until they pass the exit Standards, maybe with limited re-sits). *Overall, there is evidence in these contested Figures of a decline in performance in Examination related to entry standards but not in the final assessments from the training course.*

3. From my investigations it does appear that there is one feature of this debate not explored in either Saunders or Annan’s first paper (although confirmed to me via email by Saunders following discussion, noted in the second Annan paper; and confirmed to me by Michael Sutorius), and that is: when the ACER was used the standard for entry was Stanine 5 with exceptions below this, and for GRT2 the standard for entry was Stanine 4 – and the various data made available to me show this to be the case. In the Annan rebuttal, for example, the Chart on p. 19 shows the relation between all assessments and the stanines of entry – there are only two entries with ACER below Stanine 5, and many with GRT2 below Stanine 5. Given Stanine 5 and above excludes the bottom 40% and Stanine 4 and above excludes the bottom 23% of the population then *this change from Stanine 5 to Stanine 4 is, prima facie, a decline in the entry standards.*

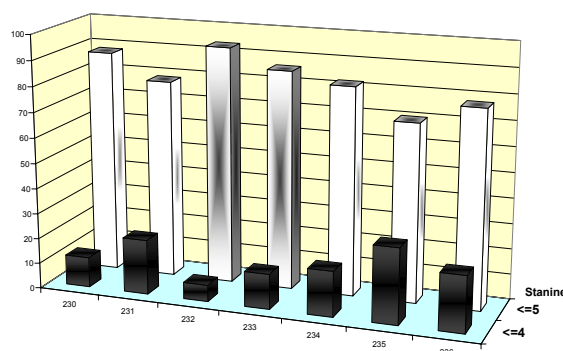
A key factor is whether the tests are measuring similar attributes (and they appear to be), and whether the norm-sample is comparable (which is likely). It may be that the scores on the ACER are inflated as there are claims that the test had been exposed and was included in some (non-Police) training programs to prepare applicants to sit the ACER test; notwithstanding this decline from an “expected” minimum of Stanine 5 to Stanine 4 does represent a substantive change in the minimum entry standard.

From the Sutorius data, the following shows the average ACER Stanine, and the GRT2 Abstract, Verbal and Numerical averages (excluding any Wing with less than 5 applicants). There is a decline in the average Stanine when moving from the ACER to the GRT2 tests.



4. Saunders claimed that given the power of reasoning tests to predict police job performance, which manifests in major differences after many years “when staff in sworn roles should be looking for supervisory roles or performing in areas of speciality” (p. 8), then any current decline in entry standards must only have a later major negative impact. This is not contested by Annan, nor would it be expected to be contested if there are no declines in the graduating standards from Police training. Saunders claim, however, needs to be moderated as there are other selection criteria alongside the reasoning tests, and there is evidence that the training program reduces the impact of differences in selection into the Police.
  
5. Saunders asks whether the police are “currently attracting ‘at least’ average candidates” (p. 9). The answer, as noted above, is that (a minimum of) Stanine 4 includes IQs of 90-96, which would be considered ‘at least’ or close to average; and given a profile of many measures across many dimensions is actually used then it is reasonable to conclude that the pool of applicants is ‘at least’ average or close to average. When Stanine 5 was used as the expected minimum from the ACER then these would include IQs from 97-104 – certainly “average”. Further Michael Sutorius provided me with a data file of 3830 recruits who started at RNZPC since July 2000 and 60% had a stanine entered in this database. The average stanine from the ACER entrants (N=997) was 6.78 which corresponds to an average IQ of 109, and from the GRT2 (N=1263) was 6.09 for Abstract Reasoning, 6.01 for Numerical Reasoning, and 6.15 for Verbal Reasoning – which corresponds with an IQ of 105, 104, and 105, respectively. *It appears that the typical Police recruit is average to above average in reasoning – and this has not changed from a move from the ACER to GRT2 and any related changes in standards.*
  
6. There is an increased growth in the bottom end of the distribution over time (Saunders, p. 12). My representation shows some but not major changes in the bottom end of the distribution with the exception of the doubling of the number of recruits in Stanines  $\leq 4$  for the last two wings (235, 236). (Note, Saunders noted to me that the data on p. 12 related only to the Verbal Reasoning derived from the Recruiting summaries [see comment on their accuracy below] and he used these given the dominant role these attributes appeared to have, and would have footnoted this had he been able to finish the report.)





7. Saunders argues that retesting has been “stretched to allow entry of marginal candidates”. The claim is that those recruits who score below the cut-score are allowed to re-sit the (same) test, and there is evidence of database entries for an applicant sitting twice in three days (and not after the recommended 6 weeks). From the data provided to me, there were 60 applicants who were allowed to re-sit. The average days between sitting was 63 days (9 weeks, with only 7 in fewer than 6 weeks and two of those on the same day - and these two did not change enough to be selected!). There were low correlations between performance on the two sittings (Abstract,  $r = .38$ , Numeric,  $r = .09$ , Verbal,  $r = .29$ ) and the average increases were small for Abstract = 2.80 and Numeric = 2.41; but higher on Verbal = 5.15. There were 9 who changed by more than 10 points on Verbal reasoning). It was the case, not surprisingly, that candidates with scores close to the cut of Stanine 3 and 4 were more likely to be re-sit, and there were 30 of these for whom the decision was to select and 30 who remained not selected as a consequence of the re-sitting policy.

A further claim by Saunders to me in discussion was that there is evidence of recruits with Stanines of 2 who were being permitted to re-sit and this is clearly contrary to current Guidelines. (He showed me files of 12 such recruits.) In the data provided to me by Michael Sutorius there certainly were cases where recruits with Stanines of 2 were allowed to re-sit. For example, one had Stanine 2 on Abstract, Stanine 2 on Numerical, and Stanine 3 on Verbal, and was permitted to re-sit to achieve Stanine 3, 4, and 3, respectively (but not then selected). Further, there is 1 candidate with Stanines of 2 for Abstract, and 2 with Stanine 1 for Numerical that are in the file of selected candidates for the Police training course. One possible explanation is that there were a group of recruits that completed both the ACER and GRT2 (as part of the evaluation of the new test) - and while a recruit may have a low Stanine on the new test, it is important to reconcile which test was used as the basis for the admission decision.

As Saunders noted, it is well worth studying those who gain selection by any re-sit method, and track “this group over an extended period” (p. 15). It is always unwise to re-test on the same set of high-stakes items too soon as practice and any teaching to the test can allow for increased scores. When tests become tasks “to be beaten” then there is a need to question the testing protocols, but as Saunders notes, where there are anomalies (e.g., high on Verbal but low on Abstract) there may well be test factors that can cause this difference – and good practice would lead to closer investigation of any discrepancies.

The inference in this discussion is that minimal entry standards are being stretched to allow less-than-ready applicants into the Police. While the numbers of these appears few indeed, there are some anomalies – and the perception from the public would certainly be that the standards are being lowered if there are “any” applicants below the prescribed standards. Given the numbers seem so few, it is recommended that minimum criteria for these Reasoning tests are tightly adhered to, and there be no

exceptions to the minimum of Stanine 4, and all those in Stanine 4 be considered by a more defined set of processes. (Also see recommendation below re Data processes which impact on this discussion.)

***Recommendation 5: It is recommended that minimum criteria for these Reasoning tests are tightly adhered to, and there be no exceptions to the minimum of Stanine 4, and all those in Stanine 4 be considered by a more defined set of processes.***

8. In response to the TenOne article by Annan (Nov 2006), Saunders comments that it is concerning that there are *current* serving officers who score in the bottom stanines (lowest 10%) in reasoning. This may, however, reflect a need for a closer inspection of professional development than a statement about selection processes. Further, as the standard error of the tests is about 4 points on the reasoning tests, then some movement up and down over time should be expected.

The ***counter argument*** is that:

9. There are an (unspecified) “number of current serving Police officers ... [who] would not be able to continue through the recruiting process today, without re-sitting the test and passing” (Annan) + “The tests have shown that on the whole, NZ police staff are capable people” – and thus the current cohorts of recruits are acceptable. Annan provided graphs of 84 current recruits who were successful on the GRT2, 255 police who had sat the old ACER test (then in place) and the new GRT2 test, and (unspecified numbers of) current Police at constable level with 5-10 years service on the GRT2 test. He concluded that across all three reasoning tests, “the more recent wings have higher average stanine scores.”

Three line graphs are presented. There is truncated variance in the recent applicants as they could not proceed unless they scored at Stanine 4 or higher; and there are no applicants in Stanine 4 for Numerical Reasoning for the GRT2 which is an error, and there are recruits who scored Stanine 2-4 from the ACER when these should have been declined (the latter may be due to using scores before re-sits although the text notes that this option was not available to these Wings). Annan (email) has indicated that there is a problem with the Numerical Reasoning; as he found some corrections needed to be made to the total sample size, and there was an error in the look-up table for converting raw to Stanines. The graph would look different but the fundamental message would not change.

The correlation between the ACER and GRT2 test is not perfect thus it should be expected that there is not complete overlap. For example, if the correlation between the ACER and GRT2, or the test-retest scores on GRT2 (or ACER) was as high as .90 then it would be expected that about 18% - 20% of candidates who scored above Stanine 4 on the ACER would, on retesting, score lower than Stanine 4 – and this is indeed the case: the movement of current serving police across the Stanines could simply be a function of measurement error.

Further, the correlation of the scores from the GRT2 and the stanine on the ACER is very high for Abstract reasoning ( $r = .96$ ) and Verbal reasoning ( $r = .98$ ) but low for Numerical Reasoning ( $r = .30$ ). This should provide much comfort in the predictiveness of the selection tools and subsequent performance (but again the usefulness of the Numerical Reasoning test may need to be questioned). Saunders

correctly notes also that there is need for caution when the sample of re-tested Police is only 65-84 from a total of 7000+.

Notwithstanding these issues, the comparisons should provide some comfort that the performance of the various groups of Police on the ACER and GRT2 tests is similar in their distribution of reasoning (except for Numerical reasoning).

10. Annan concluded that 26% of the current serving officers would have to re-sit and pass the reasoning test – if they were to “re-enter” the Police. This analysis is confounded as the current serving police are older and more experienced, did not enter under the currently used test, this re-testing was not conducted under the high stakes environment experienced by new recruits, and there are unknown issues with equating these two tests.
11. Annan provides Stanine conversions for the “general population of New Zealanders” (although the attributes of this sample are not stated). From the data presented it is clear that police applicants are higher on reasoning than their New Zealand peers – 94% of Police compared to 77% of New Zealanders score above Stanine 4 in Verbal reasoning, 89% compared to 77% in Numerical reasoning, and 83% vs. 89% in Abstract reasoning. In general, Police recruits are more proficient in reasoning than the New Zealand population.
12. There is a claim that on the GRT2 Police who would score below the cut-score for Stanine 4 “are able to achieve at University” (Annan, 2006, p. 16). The correlation of most selection tests and University success is between .20 to .40, which shows remarkable room for those who DO NOT gain selection to continue to University successfully. The issue, however, is to maximise a scarce resource (University opportunity) to those most likely (but not guaranteed) to succeed. I have no doubt that some candidates with Stanine 3 or less on the GRT2 could also succeed at University – but which kind of University (as there are now many tertiary institutions in New Zealand with much variance in quality), there are numerous degrees and certificates, and University success can be defined in multiple ways. Saunders is more likely to be correct with the claim that a person scoring low on the GRT2 (and I would add the ACER tests) will be unlikely to possess the skill required for University study (my emphasis).

I was provided the data on 997 Police who entered via the ACER and 1262 who entered via the GRT2, and their results from the Victoria University paper LEGL114 Introduction to Criminal Law. I converted all results in this course to Pass (A, B, C and those credited Pass) and Fail) leading to a creditable 92% pass rate. There are no differences in pass rates related to the ACER Stanine (effect-size = .02), or on the Abstract GRT2 (-.03). Police with higher Verbal reasoning ( $d=.21$ ) and lower Numerical reasoning ( $-.27$ ) are more likely to pass. The pass rates in this Law paper have dropped over the past five years (based on starting date in the training course: 2003 98%, 2004 95%, 2005 90%, 2006 90%, and 2007 85%;  $F=3.54$ ,  $df=4$ , 2070,  $p<.001$ ).

Average Stanine for those passing and failing the Law paper.

	Failed	Passed	Effect-size
<i>ACER</i>	6.72	6.74	0.02
sd	1.37	1.41	
No.	57	653	
<i>GRT2</i>			
<i>Abstract</i>	6.07	6.03	-0.03
sd	1.46	1.33	
No.	41	235	
<i>Numerical</i>	6.20	5.82	-0.27
sd	1.49	1.32	
No.	41	235	
<i>Verbal</i>	5.63	5.92	0.21
sd	1.30	1.39	
No.	41	235	

There was no statistically significant relations between the ACER Stanines and GRT2 Numerical reasoning and the Passing of this Law paper, but those higher on Abstract and particularly Verbal reasoning were more likely to pass this paper.

No. of recruits that passed or failed the Law paper related back to their stanines

	ACER		Abstract		Numerical		Verbal	
Stanine	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail
3							1	0
4	11	0	25	7	37	5	40	9
5	135	13	76	9	71	12	59	11
6	180	16	52	9	60	6	62	12
7	114	10	48	7	44	8	40	5
8	111	10	26	8	13	8	22	3
9	102	8	10	1	11	2	12	1
Total	653	57	237	41	236	41	236	41

Note, these Law passes say little about the quality of the ACER and GRT2 cut-score for entry. These analyses provide confidence in using the GRT2 Verbal and Abstract reasoning tests as reasonable predictors of success in the University Law paper.

### Overall comment on “decline” or not.

Is there reason to believe that the quality of police recruits has declined? On the basis of the change from a minimum of Stanine 5 under ACER to Stanine 4 under GRT2 this is a potential decline. Indeed, there is a decline in passing the first Written examination over the past 25 Wings, but no evidence of decline in final assessments of the training program related to this change.

It is critical to note that changing minimum cut-scores does NOT necessarily lead to claims about overall decline in standard – as a key factor is the percentage of recruits who enter at the *lower* ends compared to the *overall* distribution. There are about 10% of the cohort entering with a Stanine 4 in one of the three GRT2 tests. In the data provided to me, there are 8% who have an average across the three GRT2 tests less than Stanine 5: of these 43 recruits, 33 had at least one Stanine 5 and the other two Stanine 4; and 5 had Stanine 4 for all three tests – applicants certainly eligible under

the “Guidelines” and very likely were recruited considering other selection criteria. Only 5 had Stanines that should cause concern – (Abstract, Numerical, Verbal: 5,1,3; 4,3,4; and 3 with 5,3,5). Overall, not convincing evidence to say that a) the selection criteria have been misused, or b) that the minimum standard is a cause for major concern – certainly not sufficient to justify the current level of debates. There is no suggestion in my statements that a Stanine 4 or 5 should be the minimum cut-score – this should be a consequence of first, a job analysis, and b) a standard setting exercise – both described above.

It is important throughout these debates about the Reasoning tests, that these are NOT the only selection criteria – a fact noted in all reports (see Annan, rebuttal, p. 13 for a more extensive discussion of this process).

It has also been noted to me that when candidates are accepted into the training program they are then sworn in as “police” so the critical decision is at this entry point and not at the end of training. Most other professions are ‘sworn in’ on the successful completion of training not when entering training. Discussion of the merits of this decision is outside the brief for this paper, but does highlight the importance of getting the selection process correct.

***Overall, while there is evidence that the minimum entry criteria on the Reasoning tests has declined, there is no evidence to believe that the overall performance of graduates from the training program has changed.*** It is not clear that any decline is attributable to the change in assessment, but is more likely related to decisions about the standards/Guidelines/Rules of entry. It does appear that there is more need for tutorial and remedial assistance (particularly in light of Written Test One), but there is evidence on the final assessments indicating it may be having the expected effect. It appears that the typical Police recruit is average to above average in reasoning compared to the NZ population – and this has not changed when moving from the ACER to GRT2 and any related changes in standards. There should be tightening of the “Guidelines” for making decisions about marginal candidates and when and how frequently re-sitting is permitted. There is much evidence that using Reasoning tests is valuable as a predictor – although there is more merit in using Verbal and Abstract reasoning than Numerical Reasoning, and that Verbal Reasoning is among the more critical skills to pass the Victoria University law paper.

#### **G. The process for conducting research within the Police**

There is a high need for a debate within the Police about the processes for conducting research activities, a need for more dependable data bases, and a need for higher levels of “interpretations” from these various selection profiles.

There needs to be questions about why the current debate has been permitted to occur. When there are public comments about research issues such as those cited in this report, then it is desirable that the “research” section of the Police provide the best and most dependable source of evidence on such issues. The comments by Annan (in Ten-One November, 2006) appear to be a most worthwhile response to some earlier claims. I then understand that Saunders was asked to further investigate these claims and this led to his draft. Saunders was most adamant about the status of his report – it was an uncorrected draft, which after showing to his sponsor, he was not permitted to further explore (to correct language, check data, further explore, etc.) – see p. 10 of Annan’s rebuttal for further comments on this process, which raises serious concerns about process. Certainly the report does appear to be in draft mode. It seems that his report was then used in other forums and this led to Annan’s paper. I am not privy to the discussions about these two papers at that stage, but a more suitable process would involve getting the various report writers together to work through the evidence

(this is *not* a comment on Annan or Saunders but on the subsequent process). It is always the case in research endeavours that there can be multiple views, contested interpretations of data, and any report prepared by Police needs to follow a stipulated process. It is recommended that the Police develop a Policy for the release of any research paper including processes for internally resolving any differences in interpretations and conclusions. No Police internal report, no matter what the status, should be used by Police and certainly not in public discussion until it has been so approved via this Policy. All discussions up to the signed release should be appropriately confidential within the Police, follow a reasonable process of debates and discussions, all participants need protection when they express their interpretations within the Police, and a method for resolving disagreements decided – prior to any release to any source.

***Recommendation 6: It is recommended that the Police develop a Policy for the release of any research paper including processes for internally resolving any differences in interpretations and conclusions.***

It is also concerning that there is debate about the data sources. There seems to be multiple data sets, not all corresponding to each other, the data I was sent (acknowledging the fast turn around requested) included dirty data (e.g., Stanines > 9), and Annan would support my claim that it is “important that we continue to develop one data source rather than multiple data sources ... it is a matter of prioritisation”. There are disagreements between the PNHQ database and the recruiting summaries, many missing stanine scores in the database provided, and different sets of data used for the various analyses (e.g., Saunders used all recruits in parts of his analyses whereas Saunders used only those who graduated). The error rate in some of the data provided to me (in various forms) and also noted in the Saunders report is unacceptable. In the data files sent to me there were major errors (Stanines > 9, etc.) that, if not checked, could skew any subsequent analyses. The letters re. recruitment are also sub-standard. Saunders reports that there is a 24% error rate in these letters when reporting stanine scores – by transposing scores, typing errors, and shifts in the database re. calculation of stanine cut-scores! Some have hand-written notes about re-sits and more concerning is those not subsequently corrected if there were indeed re-sits. I cannot replicate a 24% error rate from Appendix 3 of the Saunders report – but certainly there are a large number of raw scores on the three reasoning tests which have stanines that do not a) correspond to the same conversion using the tables provided in the Annan report, and/or b) where the same raw score leads to a different stanine conversion. There are hand written notations on these letters, and certainly too many cases where there are causes for questioning the veracity of the conversions – e.g., there are four applicants who were admitted but their stanines were below the appropriate stanine cut-score. There are more that seem to have low stanines but hand written comments about re-sits. Whatever the right “error rate” there are enough questions about these data to ask for a more dependable system to record and send information about these entry test scores. I noted comments in both reports, and in my own analyses of the data provided to Recommend that a Review of the use, usefulness and dependability of the assessment data held by Police on all those who apply and/or are recruited, how these data and interpretations are dependably communicated to all who have need/ permission, and how it is archived to be of use to all who have access to these data.

***Recommendation 7: It is recommended that a Review of the use, usefulness and dependability of the assessment data held by Police on all those who apply and/or are recruited, how these data and interpretations are dependably communicated to all who have need/ permission, and how it is archived to be of use to all who have access to these data.***

It seems that there may be issues with different data being used by Annan, Saunders (and me) and storing data in separate Access files seems not sufficient. For example, on p. 18 of the Annan second paper it is noted that some of the discrepancies in the earlier two reports could

be related to different data sets: Saunders used all recruits, whereas Annan used only those that graduated. Saunders notes that it is those who did not graduate that are of major interest – such matters could readily have been resolved if there had been a process for conducting Police research (including bringing together the two authors and resolving such fundamental matters as sharing common data and assumptions; certainly a request made by Annan, Rebuttal paper, p. 10; and desired by Saunders, in communication with me). It is applauded that these data are used for research by Police, but both Saunders and Annan should be assured that they are using data that is replicable, dependable, and similar to all who access it. Many of the disputed issues in these papers could be more readily resolved with a more robust data filing system.

***Recommendation 8: It is recommended that an appropriate data base is developed for all to access and use, such that the raw scores of the items of each test are entered and Stanines are then automatically produced; and a decision matrix devised such that exceptions to the Guidelines not be permitted (e.g., allowing Stanine 2 to re-sit). Such an item/test data base would be valuable to subsequent item and test analyses, inform future Chernyshenko -type reports, and allow for the development of more defensible interpretative reports to recruits and trainers.***

### ***Concluding Comments***

Much of the debate in the various documents relates to attributes of specific tests – what do they measure?, what is the cut-score?, can they be gamed?, etc. This is not uncommon when test scores become more reified than can be defended from a psychometric viewpoint. An alternative, and the basis of my own asTTle research, is to concentrate more on the ***interpretations*** of the measures and of the cut-score. I note that in all papers provided, the authors still depend on the antiquated notions of validity residing in the tests and their scores (and the associated use of content, criterion, and predictive validity). Since Messick's (1989) significant contributions to the field of testing, validity is now conceived in terms of the interpretations that we make from tests: "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the "adequacy" and "appropriateness" of "inferences" and "actions" based on test scores or other modes of assessment" (p. 13).

More attention needs to be placed on the meaning of the interpretation of scores either side of the cut-scores. Both Police and recruits should have a clear, agreed, and defensible knowledge of what proficiencies mark those who should be from those who should not be selected. In any standard setting exercise (such as in recruiting, in deciding NCEA grades, in selection to university or not), a common understanding of the differences in proficiencies between those who are selected and those who are rejected is critical – to the credibility of the selection method, to the curricula of the program (tutoring, pre-training, post-training), and to the confidence of all that the correct decisions have been made. This would entail asking groups of Police experts to complete a standard setting exercise (I recommend the Bookmark method), choosing a cut-score on a non-banded score continuum, and then to disseminate a report on the findings of this exercise. I would encourage the Police to look at the reporting engine in the asTTle tool (and I am sure many other computerised reports would also be helpful) and consider developing their own end-report, which summarises the profile information from the various assessments in a manner that is convincing, informative, and means that the debate is about the competencies and proficiencies needed for Police work – and not so much about the individual tests, or the specific scores. Such reports can be used in profiling, and for providing recruits with defensible interpretations of the meaning of the various selection tests.

Similarly, in any test situation where there are winners and losers, it is imperative to provide psychometric reports on the qualities of all the measures used – and to make these available. It is not that all police and interested parties will necessarily read them, but it is the hallmark of professionalism of testing, allows informed debate to occur, and highlights the robustness (or not) of the measurement tools in use. The Chernyshenko project is well worth repeating every few years.

The concern in both reports is about those police recruits *at or just above* the cut-score (using stanines). There is no suggestion that the *majority* of recruits are not being selected correctly, and indeed they are likely to be excellent investments into the Police. Any good employer should remain concerned about candidates selected close to the boundary, however, and a desirable manner to resolve the current issues is to conduct a study of the success of these ‘borderline’ candidates over time. Given the concerns of restriction of range, unreliability of measurement, and not seeing how those just below the cut would have performed, there is still a strong case to monitor the performance of those just above the cut-score closely. Such a research study would provide evidence as to whether this cut-score can be maintained, or should be increased (e.g., if they exhibit less success, more stress, lower work performance than those well above the cut-score, and compared to desirable standards for such indicators). More important a more defensible standard setting model needs to be used to decide on these cut-scores.

*Two final comments.* First, if there remains doubt as to the competencies of police recruitment, it could all be out-source selection to ensure fairness but with the consequential loss of much knowledge brokering *within* the Police. Sometimes the perception of fairness demands some distance, but the costs (in terms of money and in police learning) can be high. Second, many of the above issues could be remedied:

- if there was a robust job analysis as the criterion of validity to choose the appropriate suite of selection tests;
- if there was a robust standard setting method used to assist in determining the cut-scores on these tests; and
- if interpretation of the scores (not banding for simplicity) was emphasised.



## References

- Barrett, P. (2006). *The use of standardized-normatively scaled scores within a performance-oriented selection process*. Whitepaper Series #2. Retrieved 1 August 2007 from [http://www.pbarrett.net/whitepapers/norm\\_scores\\_in\\_selection.pdf](http://www.pbarrett.net/whitepapers/norm_scores_in_selection.pdf).
- Brannick, M.T., & Levine, E. (2002). *Job analysis: Methods, research, and applications for human resource management in the new millennium*. Thousand Oaks, CA: Sage.
- Chernyshenko, S. (2005). *Report on psychometric evaluation of the General Reasoning Test (GRT2) for the New Zealand Police: Measurement equivalence across ethnic and gender groups*. Christchurch, New Zealand: Drasgow Consulting Group.
- Cizek, G. J. (Ed.) (2001). *Setting performance standards. Concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Equal Employment Opportunity Coordinating Council (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43, 38290-38315.
- Hattie, J. A., & Brown, G. T. L. (2003). Standard setting for asTTle reading: A comparison of methods (Tech. Rep. No. 21). Auckland, New Zealand: University of Auckland/Ministry of Education.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 485-514). New York: Macmillan.
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425-461.
- Kolen, M. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement*, 25(2), 97-110.
- Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K. & Patz, R. J. (2003). The Bookmark procedure: Methodology and recent implementations. Paper presented at the 1998 Annual Meeting of the National Council on Measurement in Education. San Diego, CA.
- Mariani, M. (2001, Fall). O\*NET update. *Occupational Outlook Quarterly*, 26-27. Retrieved from <http://www.onetcenter.org/>
- McEntire, L. E., Dailey, L. R., Osburn, H. K., & Mumford, M. D. (2006). Innovations in job analysis: Development and application of metrics to analyze job data. *Human Resource Management Review*, 16, 310-323.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 13-103). New York: Macmillan.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R., (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek & G. L. Cizek (Eds.). *Setting performance standards: Concepts, methods and perspectives* (pp. 249-282). Hillsdale, NJ: Lawrence Erlbaum Associates.
- O'Callaghan, T. (2003). *New Zealand Police assessment testing for recruitment and selection project*. Wellington, New Zealand: OPRA Consulting Group.
- Plake, B. S. (1995). The performance domain and the structure of the decision space. *Applied Measurement in Education*, 8, 1, 3-14.
- Schmidt, F. L. (1991). Why all banding procedures in personnel selection are logically flawed. *Human Performance*, 4, 265-277.
- Schmidt, F. L., & Hunter, J. E. (1995). The fatal contradiction in banding: Is statistical rational is logically inconsistent with its operational procedures. *Human Performance*, 8(3), 203-214.
- Shepard, L. A. (1993). Evaluation test validity. *Review of Research in Education*, 19, 405-450.
- Wilson, M. A. (2007). A history of job analysis. In L. L. Koppes (Ed.), *Historical perspective in industrial and organizational psychology* (pp. 219-241). Mahwah, NJ: Lawrence Erlbaum Associates.

## ***Appendix 1: Job Analysis for Police Patrol Officers***

(Source: <http://online.onetcenter.org/link/summary/33-3051.01>)

### **Knowledge**

**Law and Government** — Knowledge of laws, legal codes, court procedures, precedents, government regulations, executive orders, agency rules, and the democratic political process.

**Public Safety and Security** — Knowledge of relevant equipment, policies, procedures, and strategies to promote effective local, state, or national security operations for the protection of people, data, property, and institutions.

**English Language** — Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition, and grammar.

**Customer and Personal Service** — Knowledge of principles and processes for providing customer and personal services. This includes customer needs assessment, meeting quality standards for services, and evaluation of customer satisfaction.

**Education and Training** — Knowledge of principles and methods for curriculum and training design, teaching and instruction for individuals and groups, and the measurement of training effects.

**Psychology** — Knowledge of human behavior and performance; individual differences in ability, personality, and interests; learning and motivation; psychological research methods; and the assessment and treatment of behavioral and affective disorders.

**Administration and Management** — Knowledge of business and management principles involved in strategic planning, resource allocation, human resources modeling, leadership technique, production methods, and coordination of people and resources.

**Telecommunications** — Knowledge of transmission, broadcasting, switching, control, and operation of telecommunications systems.

**Clerical** — Knowledge of administrative and clerical procedures and systems such as word processing, managing files and records, stenography and transcription, designing forms, and other office procedures and terminology.

**Transportation** — Knowledge of principles and methods for moving people or goods by air, rail, sea, or road, including the relative costs and benefits.

### **Skills**

**Judgment and Decision Making** — Considering the relative costs and benefits of potential actions to choose the most appropriate one.

**Active Listening** — Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.

**Critical Thinking** — Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to problems.

**Writing** — Communicating effectively in writing as appropriate for the needs of the audience.

**Speaking** — Talking to others to convey information effectively.

**Reading Comprehension** — Understanding written sentences and paragraphs in work related documents.

**Social Perceptiveness** — Being aware of others' reactions and understanding why they react as they do.

**Negotiation** — Bringing others together and trying to reconcile differences.

**Persuasion** — Persuading others to change their minds or behavior.

**Active Learning** — Understanding the implications of new information for both current and future problem-solving and decision-making.

## Abilities

**Inductive Reasoning** — The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).

**Oral Comprehension** — The ability to listen to and understand information and ideas presented through spoken words and sentences.

**Near Vision** — The ability to see details at close range (within a few feet of the observer).

**Oral Expression** — The ability to communicate information and ideas in speaking so others will understand.

**Deductive Reasoning** — The ability to apply general rules to specific problems to produce answers that make sense.

**Far Vision** — The ability to see details at a distance.

**Problem Sensitivity** — The ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing there is a problem.

**Speech Clarity** — The ability to speak clearly so others can understand you.

**Speech Recognition** — The ability to identify and understand the speech of another person.

**Flexibility of Closure** — The ability to identify or detect a known pattern (a figure, object, word, or sound) that is hidden in other distracting material.

## Work Activities

**Getting Information** — Observing, receiving, and otherwise obtaining information from all relevant sources.

**Performing for or Working Directly with the Public** — Performing for people or dealing directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests.

**Operating Vehicles, Mechanized Devices, or Equipment** — Running, maneuvering, navigating, or driving vehicles or mechanized equipment, such as forklifts, passenger vehicles, aircraft, or water craft.

**Resolving Conflicts and Negotiating with Others** — Handling complaints, settling disputes, and resolving grievances and conflicts, or otherwise negotiating with others.

**Identifying Objects, Actions, and Events** — Identifying information by categorizing, estimating, recognizing differences or similarities, and detecting changes in circumstances or

events.

**Making Decisions and Solving Problems** — Analyzing information and evaluating results to choose the best solution and solve problems.

**Communicating with Supervisors, Peers, or Subordinates** — Providing information to supervisors, co-workers, and subordinates by telephone, in written form, e-mail, or in person.

**Communicating with Persons Outside Organization** — Communicating with people outside the organization, representing the organization to customers, the public, government, and other external sources. This information can be exchanged in person, in writing, or by telephone or e-mail.

**Evaluating Information to Determine Compliance with Standards** — Using relevant information and individual judgment to determine whether events or processes comply with laws, regulations, or standards.

**Documenting/Recording Information** — Entering, transcribing, recording, storing, or maintaining information in written or electronic/magnetic form.

## Work Context

**In an Enclosed Vehicle or Equipment** — How often does this job require working in a closed vehicle or equipment (e.g., car)?

**Face-to-Face Discussions** — How often do you have to have face-to-face discussions with individuals or teams in this job?

**Deal With External Customers** — How important is it to work with external customers or the public in this job?

**Contact With Others** — How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?

**Freedom to Make Decisions** — How much decision making freedom, without supervision, does the job offer?

**Frequency of Conflict Situations** — How often are there conflict situations the employee has to face in this job?

**Frequency of Decision Making** — How frequently is the worker required to make decisions that affect other people, the financial resources, and/or the image and reputation of the organization?

**Deal With Unpleasant or Angry People** — How frequently does the worker have to deal with unpleasant, angry, or discourteous individuals as part of the job requirements?

**Work With Work Group or Team** — How important is it to work with others in a group or team in this job?

**Impact of Decisions on Co-workers or Company Results** — How do the decisions an employee makes impact the results of co-workers, clients or the company?

## Job Zone

**Title** Job Zone Three: Medium Preparation Needed

**Overall** Previous work-related skill, knowledge, or experience is required for these

**Experience** occupations. For example, an electrician must have completed three or four years of apprenticeship or several years of vocational training, and often must have passed a licensing exam, in order to perform the job.

**Job Training** Employees in these occupations usually need one or two years of training involving both on-the-job experience and informal training with experienced workers.

**Job Zone Examples** These occupations usually involve using communication and organizational skills to coordinate, supervise, manage, or train others to accomplish goals. Examples include funeral directors, electricians, forest and conservation technicians, legal secretaries, interviewers, and insurance sales agents.

**SVP Range** (6.0 to < 7.0)

**Education** Most occupations in this zone require training in vocational schools, related on-the-job experience, or an associate's degree. Some may require a bachelor's degree.

## Interests

Interest code: **SRE**

**Social** — Social occupations frequently involve working with, communicating with, and teaching people. These occupations often involve helping or providing service to others.

**Realistic** — Realistic occupations frequently involve work activities that include practical, hands-on problems and solutions. They often deal with plants, animals, and real-world materials like wood, tools, and machinery. Many of the occupations require working outside, and do not involve a lot of paperwork or working closely with others.

**Enterprising** — Enterprising occupations frequently involve starting up and carrying out projects. These occupations can involve leading people and making many decisions. Sometimes they require risk taking and often deal with business.

**Conventional** — Conventional occupations frequently involve following set procedures and routines. These occupations can include working with data and details more than with ideas. Usually there is a clear line of authority to follow.

## Work Styles

**Integrity** — Job requires being honest and ethical.

**Self Control** — Job requires maintaining composure, keeping emotions in check, controlling anger, and avoiding aggressive behavior, even in very difficult situations.

**Stress Tolerance** — Job requires accepting criticism and dealing calmly and effectively with high stress situations.

**Attention to Detail** — Job requires being careful about detail and thorough in completing work tasks.

**Dependability** — Job requires being reliable, responsible, and dependable, and fulfilling obligations.

**Concern for Others** — Job requires being sensitive to others' needs and feelings and being understanding and helpful on the job.

**Independence** — Job requires developing one's own ways of doing things, guiding oneself with little or no supervision, and depending on oneself to get things done.

**Initiative** — Job requires a willingness to take on responsibilities and challenges.

**Cooperation** — Job requires being pleasant with others on the job and displaying a good-natured, cooperative attitude.

**Adaptability/Flexibility** — Job requires being open to change (positive or negative) and to considerable variety in the workplace.

## Work Values

**Achievement** — Occupations that satisfy this work value are results oriented and allow employees to use their strongest abilities, giving them a feeling of accomplishment. Corresponding needs are Ability Utilization and Achievement.

**Support** — Occupations that satisfy this work value offer supportive management that stands behind employees. Corresponding needs are Company Policies, Supervision: Human Relations and Supervision: Technical.

## Wages & Employment Trends

### National

Median wages data collected from **Police and Sheriff's Patrol Officers**.

Employment data collected from **Police and Sheriff's Patrol Officers**.

**Median wages (2005)** \$US22.25 hourly, \$US46,290 annual

**Employment (2004)** 639,000 employees in USA

**Projected growth (2004-2014)** ■■■Average (10-20%)

**Projected need (2004-2014)** 264,000 additional employees

## POLICE STANDARDS AND ASSESSMENT PRACTICE (TO PROBATIONARY CONSTABLE LEVEL)

### Overview of Key Points

This Report provides an independent study of the efficacy of the data, methodology and conclusions offered in the reports written by Saunders (“Recruit training and the strategic direction of Policy – Standards and performance”), and Wayne Annan (“Analysis of staff with low stanine score entering RNZPC”). There are seven major issues: The lack of any reference to a job analysis; is general mental ability a useful predictor in police work? the choice of assessment tools described in these papers; the misuse of stanines; the lack of discussion about standard setting for setting cut-scores; the question of whether standards of police recruits have declined? and processes for conducting research within the Police.

From this review, it is concluded that while there is evidence that the *minimum* entry criteria on the Reasoning tests has declined, there is no evidence to believe that the *overall* performance of graduates from the training program has changed. It is not clear that any decline is attributable to the change in assessment, but is more likely related to decisions about the standards/Guidelines/Rules of entry. It does appear that there is more need for tutorial and remedial assistance (particularly in light of Written Test One), but there is evidence on the final assessments indicating this assistance may be having the expected effect. It appears that the typical Police recruit is average to above average in reasoning compared to the NZ population – and this has not changed when moving from the ACER to GRT2 and any related changes in standards. There should be tightening of the “Guidelines” for making decisions about marginal candidates and when and how frequently re-sitting is permitted. There is much evidence that using Reasoning tests is valuable as a predictor – although there is more merit in using Verbal and Abstract Reasoning than Numerical Reasoning, and that Verbal Reasoning is among the more critical skills to pass the Victoria University law paper.

There are eight **Recommendations**:

- 1 That the Police undertake a job analysis, with particular reference to the first years of policing, to provide the basis for then choosing selection tests and desired cut-scores on those tests to enter the profession.
- 2 That there needs to be clear advice on the consequences of scoring in Stanine 1-3 (there needs to be resolution between the various “Rules” and “Guidelines” for minimal entry using the GRT2 tests).
- 3 That the raw scores or an IRT-based proficiency estimate should be used—a cut-score established, and the relevant error of measurement used to make decisions about selection or rejection.
- 4 That the Police use a standard-setting process (such as the Bookmark method) to determine the appropriate cut-score on the various assessments it is using in the selection of Police recruits.
- 5 That minimum criteria for these Reasoning tests are tightly adhered to, and there be no exceptions to the minimum of Stanine 4, and all those in Stanine 4 be considered by a more defined set of processes.
- 6 That the Police develop a Policy for the release of any research paper including processes for internally resolving any differences in interpretations and conclusions.
- 7 That a Review of the use, usefulness and dependability of the assessment data held by Police on all those who apply and/or are recruited, how these data and interpretations are dependably communicated to all who have need/ permission, and how it is archived to be of use to all who have access to these data.
- 8 That an appropriate data base is developed for all to access and use, such that the raw scores of the items of each test are entered and Stanines are then automatically produced; and a decision matrix devised such that exceptions to the Guidelines not be permitted (e.g., allowing Stanine 2 to re-sit). Such an item/test data base would be valuable to subsequent item and test analyses, inform future Chernyshenko -type reports, and allow for the development of more defensible interpretative reports to recruits and trainers.