This is the published version:


Available from Deakin Research Online:

[http://hdl.handle.net/10536/DRO/DU:30059560](http://hdl.handle.net/10536/DRO/DU:30059560)

Reproduced with the kind permission of the copyright owner.

Copyright : 2013, Emerald Group Publishing
A critical review of psychological instruments used in police officer selection

Jonathan Lough and Kathryn Von Treuer
Psychology, Deakin University, Burwood, Victoria, Australia

Abstract

Purpose – The purpose of this paper is to critically examine the instruments used in the screening process, with particular attention given to supporting research validation. Psychological screening is a well-established process used in the selection of employees across public safety industries, particularly in police settings. Screening in and screening out are both possible, with screening out being the most commonly used method. Little attention, however, has been given to evaluating the comparative validities of the instruments used.

Design/methodology/approach – This review investigates literature supporting the use of the Minnesota Multiphasic Personality Inventory (MMPI), the California Personality Inventory (CPI), the Inwald Personality Inventory (IPI), the Australian Institute of Forensic Psychology’s test battery (AIFP), and some other less researched tests. Research supporting the validity of each test is discussed.

Findings – It was found that no test possesses unequivocal research support, although the CPI and AIFP tests show promise. Most formal research into the validity of the instruments lacks appropriate experimental structure and is therefore less powerful as “evidence” of the utility of the instrument(s).

Practical implications – This research raises the notion that many current screening practices are likely to be adding minimal value to the selection process by way of using instruments that are not “cut out” for the job. This has implications for policy and practice at the recruitment stage of police employment.

Originality/value – This research provides a critical overview of the instruments and their validity studies rather than examining the general process of psychological screening. As such, it is useful to those working in selection who are facing the choice of psychological instrument. Possibilities for future research are presented, and development opportunities for a best practice instrument are discussed.

Keywords Management, Selection, Psychology, Screening, Validity

Paper type Literature review

General practice: psychological screening in paramilitary settings

Police and similar organizations around the world face the challenge of selecting applicants who will become successful on-the-job performers. Society expects police officers (and other public safety personnel, such as firefighters and correctional officers) to behave according to certain standards. The nature of policing makes the job dangerous, while also providing officers with unique authority. It is generally acknowledged that psychological screening, or profiling, as a personnel selection process for these organizations is a viable means of assessing candidates for employment. Job-person fit is particularly important in a public safety position. Use of profiling for selection in the police industry is well established, particularly in the USA (Guller, 1993, 1994a; Ho, 2001; White, 2008). The benefits of profiling include reduced staff turnover and decreased likelihood of poor job performance, which lead to considerable financial benefit (e.g. Choy, 1998). However, the instruments used for profiling, and the processes and research related to using them, need to be
valid in order for them to serve their purpose. It is the aim of this review to examine the instruments used in the screening process with the aim of identifying which instrument, if any, provides the greatest benefit for the selection process.

It should be noted that the vast majority of research in this field has been conducted with police officers and police candidates; as such, this review will focus on research from police-related studies. Research involving other jurisdictions does exist – for example, Lough et al. (2007) examined the differences between screened and unscreened correctional officers’ job performances. However, as the main body of work relates to police work, this review will be confined to the discussion of that sphere. This is supported by the argument that police-related findings may be partly generalizable to other jurisdictions (e.g. Barret et al., 1999; Brough, 2005).

**How psychological screening works**

The benefits of psychological screening that goes beyond the straightforward interview process are well established (Blau, 1994). Objective psychological assessment reduces risk from phenomena such as interviewer bias, inter-interview disagreement, and halo effects. The capacity of standardized instruments to allow benchmarking to an appropriate normative population further increases their usefulness.

The operational aspects of psychological assessment are well documented (e.g. Murphy and Davidshofer, 1994). Essentially, the process involves the administration of one or more valid psychological tests in order to assess a person’s suitability to a task or situation. The tests’ findings are evaluated by a psychologist, who looks for evidence or markers in the respondent’s profile that suggest suitability or unsuitability for a particular role. Psychological screening in an applied setting is the practical application of such assessments – the testee (an applicant to the police department) undertakes the assessment as part of their application. The resulting data and report are then used by police psychologists to assist in determining the suitability of the applicant for police duties.

Testing usually involves the administration of a formal instrument (such as the Minnesota Multiphasic Personality Inventory (MMPI)) and, in most cases, an interview (often with a psychologist; Blau, 1994). This review focusses on the first part of this process – the formal instrument. The functional goal of the screening process is an important element in psychological screening. Comparing the different methods of “screening in” (selection based on desirable traits) vs “screening out” (elimination based on undesirable traits) is an important issue. It is acknowledged (e.g. Hogg and Wilson, 1995; Meier et al., 1988) that screening in is a more difficult process.

However, it is generally suggested that both approaches should be used in tandem for optimal results in the police environment (e.g. Henson et al., 2010; Metchik, 1999). Measures focussed on detecting pathology, such as the MMPI, are more useful as screen-out mechanisms, while some personality instruments and scales – particularly those measuring desirable or positive traits – are potentially suited to screening in (e.g. Guller, 2003).

The timing of psychological assessments in a potential employee’s application process is an important variable that can differ according to national laws. For example, in the USA, the introduction of the Americans with Disabilities Act in 1992 effectively eliminated pure pre-employment screening using tests of psychopathology (such as the MMPI). This has resulted in these tests becoming administrable only once a conditional offer of employment has been made. This has hindered research into the effectiveness of many instruments that use American police samples, and has reduced the return on investment (ROI) benefits conferred by testing (Weiss and Inwald, 2010).

By contrast, no such legal requirement exists in Australia, where psychological screening is generally considered a standard component of the pre-employment selection process.

The usefulness of an instrument in the screening process ultimately focusses on one central question: how well does the instrument predict performance – particularly poor performance? The answer to this question lies in the research surrounding the instrument. This review will now focus on the most frequently used instruments, and will critically examine the research underpinning their use in applied settings. The reader will note that single instruments, such as the MMPI, are
compared with test batteries such as the Matrix-Predictive Uniform Law Enforcement Selection Evaluation (M-PULSE) and Australian Institute of Forensic Psychology (AIFP) system. While this may seem unfair on the surface—a battery, or suite of tests, should logically cover more ground than a single instrument—the comparison is valid on the basis of current practices. That is, many police departments use single instruments such as the MMPI as the sole psychological screening tool (Aamodt, 2010); it is therefore fair to compare single instruments with test batteries.

The MMPI
Instrument overview
The MMPI has been used in more police-related selection research than any other instrument (Aamodt, 2004). The original MMPI was designed in 1940 as a measure of psychological maladjustment for assessing patients seeking psychiatric help (Butcher, 2005). The test was revised in 1989 and re-issued as the MMPI-2. Both versions consist of 567 items to which the respondent is required to indicate either “true” or “false.” The instrument’s ten clinical scales measure various types of pathology. Any “T-score” (standardized score) above 70 (or above 65 in the MMPI-2) indicates the likelihood of psychological maladjustment. Three other non-clinical scales indicate how honest the subject was in answering the questions. These scales seek to indicate if the respondent was attempting to appear better adjusted than he or she actually was (faking good); was manufacturing or exaggerating symptoms (faking bad, or malingering); or was responding candidly.

More recently, a number of Restructured Clinical (RC) scales were developed to correct for some overlap in the original clinical scales (Graham, 2006). This was followed by the MMPI-2 Restructured Form (RF). Consisting of 338 items, this is described as more statistically rigorous and is considered to be based on well-established theoretical models of personality and pathology (Ben-Porath and Tellegen, 2008). At this stage, the RC scales and the MMPI-2 RF have had minimal research attention in the field of police applicant screening (Sellbom et al., 2007).

Early reviews of MMPI research
Burbeck and Furnham (1985) reviewed five studies on the MMPI and police selection conducted between 1964 and 1980. They concluded that “it has not been possible [y] to show whether psychological testing [the MMPI] can discriminate between police officers and members of the public, or between good and bad police officers [y] psychological testing may be useful for selecting out people suffering from some mental abnormality.” Similarly, in reviewing 18 studies on the MMPI, Simmers et al. (2003, p. 287) noted that the MMPI was originally designed to predict psychopathology and subsequently “does not appear to be a useful gauge for identifying certain behaviors in the less pathological range.” They argued that many skills that are essential for success as a police officer may go undetected.

A meta-analysis that included over 300 studies of MMPI research data by Aamodt (2004) observed that correlations between MMPI scales and measures of academy and patrol performance were low, and that the majority were not statistically significant. Correlations between MMPI scales and discipline problems and commendations also showed no correlation. Aamodt concluded that “because the F scale is comprised of items from the other clinical scales and because it is significantly related to both academy performance (R=0.11) and supervisor ratings of performance (R=-0.09), it is probably the most useful individual MMPI scale” (p. 94). This conclusion says a great deal about the utility of the other scales. While these correlations may be statistically significant, they are very small—the F-scale accounts for less than 1 percent of the variance of supervisor ratings of performance. Thus, the scales are unlikely to be of any clinical or practical assistance.

More detailed examination of MMPI research
While Aamodt’s (2004) meta-analysis provided a relatively recent technical overview of MMPI data, a more qualitative review of MMPI research, itself, is warranted. Weiss and Weiss (2010a) discussed the general state of MMPI research, and concluded that the MMPI has much to offer to the
discipline of police psychology. However, they stated that the MMPI is “not recommended as a stand-alone for screening, but it can be used in conjunction with other one or more other instruments” (p. 69). This statement appears congruent with the bulk of published literature. This review will now discuss a representative sample of MMPI-related research. A completely comprehensive discussion of all relevant studies, given the size of the literature base, is beyond the scope of this paper.

Inconclusive studies. Many studies examining the efficacy of the MMPI as a selection tool have yielded negative or inconclusive findings. Azen et al. (1973) studied 100 newly hired sheriffs in Los Angeles, and concluded that MMPI scores were unsuccessful in predicting dropouts. Research by Schoenfeld et al. (1980) and Merian et al. (1980) compared groups of problem police performers with a control group. No significant differences between the MMPI profiles of the two groups were detected in either study. Dralle and Baybrook (1985) used a similar approach and reported similar results – no significant MMPI scale correlations were present among any performance criteria for the sample of 356 police applicants.

Inwald (1988) used 16 predictor variables, including a variety of psychological tests, as well as clinical judgments studied in varying combinations. These were applied to 219 “public safety officers” who had remained on the job for five years. Among that group, 28 were fired for cause. In this case, clinical interpretation of the MMPI was the least accurate predictor of performance issues. A computer-generated prediction of termination or retention based on earlier MMPI research (Shusman et al., 1984) misidentified the greatest percentage of apparently successful officers by predicting they would be terminated. This method incorrectly rejected 36 percent of the group of 219 (78 individuals) – a Type One error.

Other studies have focussed on more specific aspects of the performance of MMPI scales. Hargrave et al. (1988) found no predictive utility on a derived aggression index, as developed from three MMPI scales. Similar findings from a commensurate methodology were reported by Castora et al. (2003). Costello et al. (1996) examined disciplinary suspension days among serving police officers, and found a 0.223 correlation with the F+Pd+Ma scale. They concluded that the relationship “appears unimpressive” (p. 302) and that further construct validation is necessary to determine the potential usefulness of the index.

Daniels and King (2002) noted that MMPI scales were unable to make any distinction between successful and unsuccessful performers, as determined by supervisor ratings. Bartol et al. (1992) reported that the MMPI was not a strong predictor of self-reported stress. By using performance ratings and MMPI to compare applicants in two major US cities, Winters (1990) concluded that there was no correlation between tests scores and job performance. The same conclusion was reached by Wright et al. (1990), who studied police performance at the end of a 14-week training course. Surrette et al. (2004) came to a similar conclusion. They examined MMPI special scale scores and supervisor ratings of 129 experienced police officers, and declared that “scores on the Good Cop/Bad Cop, Husemann Index, Goldberg Index, and Gonder Index were not significantly related to supervisor ratings” (p. 71).

Studies with positive findings. A large number of studies have reported positive data regarding the MMPI’s ability to add predictive power to police selection. However, many of these studies had design issues that undermined any research evidence of the MMPI’s utility. This section presents a representative selection of these studies.

Several authors (e.g. Bartol, 1982; Bernstein et al., 1982) reported significant findings that were based on statistical calculations of raw scores. This study method can be misleading because, in the actual practice of MMPI interpretation, raw scores are never used. Interpretation of the test results is always undertaken from the scaled scores, as the body of validity research supporting the MMPI is all based on scaled – not raw – scores (Butcher, 2005). Further, in these studies, when the reported raw scores were converted to T-scores, the statistically significant differences disappeared, and the mean T-scores were all in the normal range. This made it impossible to obtain any clinical differences. Bernstein et al. (1982) reported follow up data on 120 police who completed the MMPI
prior to being employed. Significant correlations between MMPI raw scores and field performance were reported. However, when the raw scores were converted to T-scores, all were in the average range. Under these conditions, it is unlikely that a psychologist using the test for screening applicants would find any meaningful data.

Bartol (1982) administered the MMPI to 102 male police applicants and, using raw scores, reported a statistically significant “elevation” on five MMPI scales for those officers rated “below average” through supervisor performance ratings. When the raw scores for each of the three groups of police were translated to T-scores, all were in the average range, which eliminated the ability to differentiate between the groups. This suggests that all of the applicants looked “normal” according to the usual criteria of test interpretation.

Similar issues were present in Bartol’s (1991) investigation of 600 police officers. Bartol found that officers who were “elevated” on the Pd (Psychopathic Deviate), Ma (Hypomania), or L (Lie) scales were later described as immature by supervisors.

Combining these three scores resulted in an immaturity index; however, there was no description of what constituted an elevated score. The difference in immaturity scores between those who were terminated or forced to resign and those who retained employment was described as highly significant. However, when the reported raw scores were converted to T-scores, the differences between the two groups disappeared, and all subjects scored within the average range. Therefore, the index appears to have little or no clinical utility.

Several studies (e.g. Beutler et al., 1985; Hiatt and Hargrave, 1988a b) reported statistically significant intergroup differences between one or more scores on MMPI scales, with the suggestion that such findings show the MMPI can discriminate between good and poor performers. Upon inspection, it was noted that two issues recurred in these studies. First, the differences in scale scores were so small that they would be unhelpful in the actual task of screening police applicants.

Second, despite the differences, the scores for good and poor performers were entirely within the normal range. Saxe and Reiser (1976) acknowledged this latter difficulty in practical application. They reported significant differences between several groups, but added that “these differences are all within the “normal” range and are too small [y] to have meaningful utility in clinical differentiation of successful and unsuccessful police applicants” (p. 424).

Some studies (e.g. Hooke and Krauss, 1971; Blau et al., 1993; Brewster and Stoloff, 1999) reported positive findings for the MMPI and police selection by using samples of already employed officers. The problem with this approach was that those who had been hired represented a very homogeneous group – by definition, the study subjects should have been “the best of the best.” Furthermore, the longer these subjects had been employed as police officers, the less likely it was that they represented new applicants. Research that uses this method has utility only if the findings are replicated with a study of police applicants.

Sellbom et al. (2007) acknowledged the self-limiting nature of using an already employed police sample, and subsequently corrected their analysis to compensate for this range restriction. The resulting analysis indicated moderate predictive ability of certain MMPI scales (RC scales), and lower than normal T-score elevations. A similar method was adopted by Caillouet et al. (2010) in their examination of the PSY-5 scales and facets. A dichotomous outcome variable (ceased employment vs still employed) was used as the predicted measure. Modest predictive power was found, albeit with caveats (range restrictions and scores modified by impression management levels), and the authors acknowledged that their model’s findings did “not represent a clearly significant advance for predicting law enforcement officer performance” (p. 236).

MMPI research has also used Police Academy performance as an outcome measure. Problems with this approach have been noted by Kleiman and Gordon (1988), who observed that this research relies on training performance as a validation measure. The authors discussed a large number of relevant legal cases, and noted that virtually all cases judged that there was “insufficient evidence to substantiate the validity of the selection device” (p. 93). Burkhart (1980) agreed, noting that “training academy data cannot be considered an ideal choice for criteria because such data do
not involve the officers’ responses to the real job demands” (p. 123). A number of studies based upon MMPI and training performance faced the issue described above (e.g. Gonder and Gilmore, 2004; Gough, 1950; Hargrave et al., 1986; Inwald and Shusman, 1984). Macintyre et al. (2001) used a sophisticated method to develop a framework to predict problem performers among police officers. A scale-combination and recalculation model was constructed from existing MMPI data. The sample included two matched groups of male officers. One group consisted of problematic performers, and the other group had no indications of poor performance. The study’s predictive model was more successful in identifying problem performers than good performers, despite the fact that all cases were classified by that dichotomy. The model appeared to find “added value” residing in the MMPI data, which could be useful at the personnel level. In practical terms, this would require submitting each MMPI profile to a relatively complex model developed from a few scales. While cumbersome, this method has promise for research with police applicants.

**The California Psychological Inventory (CPI) and the Inwald Personality Inventory (IPI)**

**CPI: instrument overview and review of research**
The CPI shares almost half (194) of its 434 items with the MMPI. However, this is focussed on common personality factors, rather than psychopathology. Items are true or false, and the instrument is scored on 18 scales (three of which are test-taking validity scales). Due to its shared heritage with the MMPI, and a general predisposition among practitioners to prefer the MMPI, much CPI research tends to be mentioned alongside MMPI studies. This has resulted in CPI and its potential contribution to police selection receiving diluted attention. In general, CPI-related research mimics the body of research using the MMPI; some researchers report significant relationships between CPI scales and aspects of police performance, while other researchers do not. Pugh (1985) and Hogan (1971) both reported that several of the CPI scales were significantly correlated with a variety of performance measures, including evaluations by field commanders. However, Hargrave (1985) and Sarchione et al. (1998) were not successful in linking CPI scales with aspects of police performance. Aamodt’s (2004, 2010) landmark meta-analysis of police psychology selection data showed some support for the CPI’s predictive validity. Aamodt’s analysis also revealed that a number of scales were predictive of academy performance. These scales included Intellectual Efficiency and Capacity for Status. However, the tolerance scale was particularly predictive. Aamodt noted that the tolerance scale “seems to provide the best combination of predicting supervisor ratings of performance, disciplinary problems, and academy performance” (Aamodt, 2010, p. 240).

**IPI: instrument overview and review of research**
The IPI is a 310-item (true or false) inventory designed to assess a range of behavioural and personality characteristics. The inventory has a focus on admitted behaviour patterns, such as trouble with the law, and drug and alcohol use. Similar to the MMPI, it also assesses other personality characteristics and clinical indicators of psychopathology (Inwald, 2010). Research support for the predictive validity of the IPI as a police selection tool is limited. Inwald (2010) asserted that the IPI is “a benchmark for newer tests in the selection field [y] the IPI has demonstrated that it identifies antisocial behavior patterns and is a useful predictor of police and public safety officer job performance” (p. 91). Positive findings have been reported by Detrick and Chibnall (2002) who stated that some of the IPI scales – notably, family conflicts, guardedness, and driving violations – were predictive of on-the-job performance at the one-year evaluation point. Similarly, Inwald (1988) reported that the IPI was accurate in identifying which applicants (after one year of performance) were likely to be terminated. However, a high false positive rate (36 percent) – a Type One error – was also noted. Other researchers have reported mixed findings with the IPI. Cortina et al. (1992) administered the IPI (and the MMPI) to newly hired police officers. Ignoring the problem of range restriction (that is,
they assumed that successful applicants were the best of the best), the IPI, although adequately measuring the “Big Five” personality dimensions, added little to no incremental validity (above the entry exams) for a range of performance measures, including academy performance and turnover. Similarly, Mufson and Mufson (1998) observed that four of the IPI’s scales were predictive of poor performance ratings and eventual termination; however, a small sample size (n=33) and the range restriction again limited the study’s generalizability.

Other published tests
A small group of other psychological instruments have also been applied to the police selection environment. The Personalty Assessment Inventory (PAI) has been used in a number of studies, with varying degrees of success. Weiss and Weiss (2010b) discussed the PAI in detail, and reported that, while some scales, such as the Interpersonal and Treatment scales showed “considerable potential” (p. 87), much research remains to be done, particularly concerning the criterion validity of some of the clinical scales. Moreover, the PAI is fundamentally a test of psychopathology, and has many similarities to the MMPI, which is much more widely used (Super, 2006). As with the MMPI, there remains the issue of using a test of psychopathology instead of assessing “normal” or non-clinical aspects of personality that may have a significant effect on job performance. The M-PULSE is a more recently developed instrument, which was first introduced in 2008 (Davis such as the MMPI and 16PF, and assesses a range of attitudes and beliefs specifically related to police employment duties. Initial research suggests that this test may have promise for police selection purposes (Davis and Rostow, 2010); however, much more research is required before this instrument can be considered empirically validated.

The AIFP test battery
Instrument overview
The AIFP profiling system is a collection of instruments that have been combined and administered by AIFP since the early 1990s. Its ancestor – the IFP profiling system – has been used with success in the USA, with some success supported by research (Guller, 1994b; Guller, 2003). The AIFP system consists of six separate psychological tests:
  . Candidate and Officer Personnel Survey (COPS): a test designed by AIFP, via research, for the specific purpose of selecting public safety officer applicants. This component is the flagship of the test battery. Scales include, but are not limited to, factors such as legal difficulties, alcohol and drug use, aggression, and impulsivity.
  . Edwards Personal Preference Schedule (EPPS): a broad measure of personality that assesses characteristics directly related to police performance. Scales include factors such as aggression, need to dominate others, capacity to learn from mistakes, ability to fit into a team, and ability to tolerate routine tasks.
  . Locus of Control (LOC): a measure of the extent to which a person feels in control of his or her life. This is a de facto measure of maturity.
  . How Supervise Scale (HS): a measure of judgment in interpersonal situations, and awareness of behavioral sensitivity when dealing with others.
  . Shipley Institute of Living Scale: a standardized, well-researched test of general intelligence.
  . The Opinion Survey: a measure that assesses attitudes toward enforcing the law, ranging from being rigid and unyielding to being soft and naive.

The full test consists of 540 items, with a mix of true or false items, forced choice items, and Likert scale items. It is intended to be used as a screen-out tool.

Review of AIFP research
Smyth and Byrne (1994) applied the AIFP profiling system to 51 Victorian Police recruits on their first day at the Victorian Training Academy, and found that the AIFP predictions were significantly
inaccurate in predicting trainees’ performance categories. They also found that, over time, the predictions were significantly accurate in determining recruits’ sick leave and their likelihood of being involved in investigations by the Internal Investigations Unit (IIU). That is, prior to commencing training, the system was able to successfully predict and distinguish between good and poor performers, high and low sick leave users, and recruits’ high and low likelihoods of being involved in IIU investigations.

More recently, Lough and Ryan (2005, 2006) examined the performance of AIFP screened and unscreened police constables after one year of service. The screened group demonstrably outperformed the unscreened group across most performance variables. The screened group took 11 percent less sick leave, had 58 percent less injury claims, were 84 percent less likely to have been in a motor vehicle accident while driving a police car, and were more than three times less likely to incur complaints from the general public.

Lough and Ryan (2010) further extended their examination of police cohorts over a period of three years. The AIFP screened group consistently out-performed the unscreened group; the AIFP group took significantly less sick days, had significantly fewer non-stress (injury) claims, and had significantly less motor vehicle accidents. Additionally, the screened group had a lower dropout rate, and were less likely to make serious on-the-job errors that resulted in formal disciplinary action or investigation. However, as with the earlier studies, individual officer performance ratings (e.g. supervisor reports) were not taken into account.

Other research supporting the validity of the AIFP profiling system in Australian conditions has been conducted in the corrections context. Choy (1998) compared New South Wales (NSW) corrections officers screened for selection by the AIFP battery with an unscreened group, and reported significantly lower rates of attrition and sick leave usage among the AIFP screened cohort after two years of service. Choy estimated that the NSW corrections department sick leave cost had reduced by 39 percent as a result of the screening. Similar findings were reported among Queensland (Byrne, 2001) and South Australian corrections officers (Casey et al., 2001), with screened officers taking approximately half the number of sick days of their unscreened counterparts during their first two years of employment. ROI calculations upon Casey et al.’s data revealed a monetary saving of over $850,000 for an initial outlay of $22,500 – an ROI figure of 3,815 percent.

Lough et al. (2007) tracked the sick leave of two cohorts of correctional officers (AIFP screened and unscreened) over a four-year period. They found that the screened group took significantly less sick leave in each of the first two years, and 30 percent less in total across the four years of the study. They also noted a significantly lower dropout (terminations plus resignations) among the AIFP screened group. The authors concluded that “psychological profiling as a method of selecting correctional officers appears to be a worthwhile and valid course of action” (Lough et al., 2007, p. 6).

Byrne et al. (2002) summative investigation of AIFP testing within the corrections context reinforced the utility of the process and its associated financial benefits. For example, they found that a deeper analysis of Choy’s (1998) data revealed that the screening process yielded a return on investment of approximately 3,000 percent. In addition to monetary benefits, there are other benefits that are difficult to quantify – those derived from hiring the most suitable candidates for the job, which improves staff morale and general efficiency, and reduces the likelihood of future legal action.

Further examination of AIFP research

The research evidence for the AIFP test differs from a great deal of research of other tests, particularly the MMPI, because of the frequent presence of control groups that allow for direct comparison between selected and non-selected cohorts. In most previous studies, the control groups were hired immediately prior to the introduction of the psychological screening program (and, therefore, the commencement of the experimental group), which tended to mitigate any potential maturation-based research artifacts. This approach is particularly strong for examining the effect of the test as a whole, and also allows for longitudinal observation.
Although research supporting the general validity of the AIFP appears sound, there is no published evidence of any incremental validity. None of the individual components, or any of the subscales within those components, have been linked or correlated with performance outcomes. Such research would be valuable given that the majority of scale-based research for the MMPI and CPI tends to diverge on the predictive validity of individual scales (although Aamodt’s, 2004 finding that the tolerance scale of the CPI is a valid predictor is noted).

In addition to the above, one other factor confounds the AIFP findings. Research to date has been drawn from both the AIFP (Australian) and IFP (American) data. While the two instruments have a lot of common content – including the COPS test – they are not the same instrument. Research evidence supporting the IFP test instrument does not necessarily apply to the potential validation of the AIFP test instrument, and vice versa. Future research needs to concentrate its efforts on a particular iteration of the test battery.

Discussion and development opportunity
From the material presented in this paper, it is apparent that no best practice instrument for the selection of police officers exists at this time. Each of the instruments examined have issues in their supporting research. The MMPI research is equivocal at best, research with the CPI and IPI is limited, and the AIFP research lacks scale-level incremental validity. Other instruments, such as the PAI and M-PULSE, are simply not used often enough to generate the research required to underpin their endorsement.

Therefore, it appears that there is opportunity for the development of a best practice instrument. Research devoted to this development needs to consider the factors that ensure the utility and validity of the instrument’s creation and testing. Future research should ensure:

- the development of an instrument that can assess as wide a range of attitudes and behaviors as is reasonable, rather than being focussed solely on pathologies;
- a sufficient sample size (>100);
- that appropriate outcome variables, such as on-the-job performance, are obtained in addition to performance during training (e.g. academy performance);
- that a control group or matched cohort is compared to any group selected using a (new) instrument; and
- that all components of an instrument, down to the scale level, are adequately investigated and examined for validation purposes.

Other criteria, such as the examination of non-successful applicants’ test scores, should also be considered (although this may not be feasible in American jurisdictions due to the Americans with Disabilities Act). Ultimately, the ideal instrument would be a test that is purpose-built for police selection, rather than a test that has been adapted for use in the police environment.

If an instrument were to be selected as a starting point for the development of a best practice instrument, the AIFP test battery appears to be a sound option. The AIFP instrument fulfills most of the above criteria, with the exception of availability of incremental validation data. Given the current author’s experience with the AIFP instrument, it is recommended that a thorough forensic examination of that test, with the aim of building a psychometrically sound instrument, would be a logical launching pad for the development of a best practice instrument.

References
Byrne, K. (2001), Research into the Effectiveness of the AIFP Critical Character Assessment System for Screening New Custodial Correctional Officers, AIFP Australia, Clifton Hill.


Guller, M. (2003), Predicting Performance of Law Enforcement Personnel using the Candidate and Officer Personnel Survey and other Psychological Measures, unpublished doctoral dissertation, Seton Hall University, South Orange, NJ.


About the authors
Jonathan Lough is a PhD Candidate at the Deakin University, Melbourne, Australia. He is also the Principal of the Lough Research Services, and a Psychologist and an Independent Research Consultant in Melbourne, Australia. He specializes in research design, analysis and statistics, survey writing, and psychological assessment. Over the last ten years, he has worked with a range of private and government clients, including ANZ Bank, Clemenger BBDO, and the Australian Institute of Forensic Psychology. Prior to establishing Lough
Research Services, Jonathan was a Research Psychologist at the Monash University, working in the area of public safety and injury prevention. Jonathan Lough is the corresponding author and can be contacted at: jlough@optusnet.com.au

Dr Kathryn von Treuer is the Associate Head of the School (Partnerships and Development) and the Course Chair for the Organisational Psychology Program at the Deakin University. She utilizes 20 years of industry experience in organisational psychology and senior management to inform her research, teaching and service roles at the University. She is highly regarded and sought-after by industry groups, and is a national leader in areas of organisational psychology practice and research. Dr Kathryn von Treuer completed postgraduate studies in psychology, business and education. Her doctoral dissertation was awarded the prestigious “Research Excellence” award from the Swinburne University in 2006, for investigating organisational factors that facilitate workplace innovation. She was recently invited to provide a keynote address on organisational resilience, and is the Managing Editor of the Australian and New Zealand Journal of Organisational Psychology.