

**Military Police Selection in Canada:
An evaluation of the Canadian Forces Military Police Assessment Centre (MPAC)**

by

Kathryn E. Hodgson

**A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial fulfillment of the Requirements for
the Degree of Master of Science in Applied Psychology
(Industrial/Organizational)**

July 2006, Halifax, Nova Scotia

© Kathryn E. Hodgson, 2006

**Approved: Arla L. Day, Ph.D.
Supervisor**

**Approved: Victor M. Catano, Ph.D.
Committee Member**

**Approved: James E. Cameron, Ph.D.
Committee Member**

**Approved: John V. Pumphrey, M.Sc.
External Examiner**

Date: July 24, 2006



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-25519-3

Our file Notre référence

ISBN: 978-0-494-25519-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Table of Contents	
Table of Contents	ii
Appendices.....	iii
List of Tables	iv
Acknowledgements.....	v
Introduction.....	1
Assessment Centres	2
Cognitive Ability and Job Performance.....	7
Assessment and Selection for Entry-Level Jobs.....	9
Canadian Forces Military Police Selection.....	10
CF Military Police Job Performance – The Criterion Dilemma.....	13
Overview and Research Goals.....	15
Method	16
Sample	16
Measures	17
Procedures.....	19
Results.....	23
Reliability and Factor Structure of MPAC	23
Predictive Validity	29
Discussion.....	31
Limitations, Future Research, and Implications for the CF.....	34
Specific Recommendations for the CF MP Branch.....	42
Conclusion	43
References.....	45

Appendices

A:	MPAC Competency Measures.....	55
B:	MPAC Pre-Selection Scoring Guide.....	56

List of Tables

1.	Military Police Assessment Centre Competencies.....	18
2.	MP QL3 Tested Performance Domains.....	20
3.	Internal Consistency Coefficients and Lowest Item-Total Correlations for the MPAC's 12 Competency Scales.....	23
4.	Internal Consistency Coefficients and Lowest Item-Total Correlations for the MPAC's Five Method Scales.....	24
5.	Pattern Matrix of PCA Four Factor Extraction Direct Oblimin Rotation.....	26
6.	Correlations Among MPAC Methods.....	27
7.	Mean and Standard Deviation of the Convergent Validity Coefficients for each of the 12 MPAC Competencies.....	27
8.	Mean Heterotrait-Monomethod Discriminant Validity Coefficients for each of the 6 MPAC Methods.....	28
9.	Correlations Among QL3 Performance Tests.....	29
10.	Pearson Correlation Coefficients of MPAC Factors/CFAT and QL3 Performance Tests.....	30

Acknowledgements

“...How do you thank someone who has taken you from crayons to perfume? It isn’t easy but I’ll try” (Lulu, 1967).

I owe a debt of gratitude to many people who have supported me in this academic quest. First, it would be hard for me to picture my time at Saint Mary’s University or this document without the enduring advice and guidance of Dr. Arla Day. I am sincerely humbled by, and grateful for Dr. Day’s personal commitment to seeing me through to the end of this project. I also have the utmost admiration and respect for the entire psychology department faculty who dedicate an incredible commitment of time and energy to providing quality education and research. I would particularly like to thank Dr. Vic Catano, Dr. James Cameron, Dr. Lori Francis, Dr. Camilla Holmvall, Dr. Debra Gilin, and Dr. Richard Allon for their guidance, teachings, counsel, professionalism, and patience. I would also like to extend an extra thank you to my thesis committee members who invested a lot of extra time and effort. And to April Scheller, thank you for your humour, your smiles, and for making the department run like a well-oiled machine; you make our jobs easier.

This academic quest would also not have been possible without numerous personnel of the Department of National Defence. Support and encouragement from colleagues and friends within the Personnel Selection branch and from the Military Police and Administration branches was abundant. I would particularly like to thank Vice-Admiral (Retired) Ron Buck, whose persistent prodding made me finally believe that attaining this goal was possible.

Finally, I must express my most sincere and heartfelt gratitude to my family, friends, and loved ones who, from near or far, have behind me in every step of the way: Without you, I would not have had this happy ending. Thank you for your patience, your belief in me, and most of all, for your love and support.

Abstract

Military Police Selection in Canada: An evaluation of the Canadian Forces Military Police Assessment Centre (MPAC)

By Kathryn E. Hodgson

In this study, I investigated the psychometric properties of the Military Police Assessment Centre (MPAC) including the incremental predictive ability of job performance by the MPAC beyond the predictive validity of the Canadian Forces Aptitude Test (CFAT), a test of cognitive ability. MPAC data from a sample of Canadian Forces (CF) Military Police (MP) candidates (N=323) were examined using reliability analysis, principle components analysis (PCA), and multitrait-multimethod (MTMM) to assess the reliabilities of the 12 MPAC competencies and six method scales, the MPACs factor structure, and for evidence of convergent and discriminant validity. Job performance data from a sample (N=209) and CFAT data of a sample (N=203) of the 323 candidates were examined using correlation analyses to assess the predictive validity of the MPAC and CFAT. Results of the MPAC reliability analysis indicated that the methods demonstrated higher reliabilities than the competencies. Results of the PCA and MTMM indicated that the MPAC lacked construct validity. The MPAC and the CFAT demonstrated no correlation with the job performance measure. Consequently, the incremental predictive validity of the MPAC beyond the CFAT was not assessed. Recommendations for the use and/or development of a better job performance measure to assess the criterion-related validity of the MPAC are discussed, as are limitations and directions for future research.

24 July, 2006

Military police selection in Canada:

An evaluation of the Canadian Forces Military Police Assessment Centre (MPAC)

Ineffective selection practices can lead to a multitude of organizational and legal problems (Hacker, 1996), and in extreme cases, to an organization's failure (Catano, Wiesner, Hackett, & Methot, 2005). Conversely, effective selection processes have been shown to contribute to organizational productivity and success (e.g., d'Arcimoles, 1997; Huselid, 1995; Koch & Gunter-McGrath, 1996). To be effective (i.e., to select the candidates most likely to be successful on the job, and to meet governmental regulations), a selection process must be valid, reliable, and legally defensible (Catano et al., 2005). In general, many assessment tools, such as cognitive ability tests, structured interviews, in-basket exercises, psychomotor ability tests, personality tests, and assessment centres (ACs), tend to meet these three prerequisites (Gatewood & Feild, 2001). However, the degree to which ACs are valid, reliable, and legally defensible as selection tools is dependent upon the situation and manner in which they are applied.

Different assessment instruments have demonstrated their validity when applied in appropriate circumstances. For example, cognitive ability tests have been shown to predict job performance (Hunter & Hunter, 1984) and are widely used as a selection tool for entry-level jobs (Catano et al., 2005). ACs have been found to be a good predictor of both managerial and entry-level job performance (e.g., Campbell & Bray, 1993; Cascio & Silbey, 1979; Dayan, Kasten, & Fox, 2002; Gaugler, Rosenthal, Thornton III, & Bentson, 1987; Klimoski & Brickner, 1987; Sackett, 1987). However, given that an AC is an expensive tool, Cascio and Silbey (1979) and Hinrichs (1978) argued for using straightforward and less expensive methods. Their argument is strengthened by Schmidt

and Hunter's (1998) finding that for most jobs, cognitive ability is the most important trait determinant of job and training performance. Therefore, why should organizations use an expensive AC for selection purposes rather than a cheaper cognitive ability test that may be equally as effective? This question is particularly pertinent to organizations with large and continuous recruitment programs, such as the Canadian Forces (CF). Therefore, I examined the psychometric properties of the CF's Military Police Assessment Centre (MPAC) and its effectiveness in terms of its incremental predictive validity above and beyond the predictive validity of cognitive ability.

Assessment Centres

The assessment centre (AC) is a process used to measure a pre-determined set of job-related competencies in groups of individuals (Gatewood & Field, 2001). Typically, an AC is comprised of a collection of structured assessment instruments such as interviews, simulation exercises, and leaderless group discussions, all of which are measured using multiple trained assessors (Catano et al., 2005). However, the actual construction and content of ACs varies depending on situational and organizational necessity.

The creation and application of an AC should follow a set protocol as delineated by the International Task Force on Assessment Center Guidelines (2000). First, a job analysis should be conducted to determine job-relevant behaviours, which are then classified into competencies that can be assessed. Once the competencies are identified, reliable, valid, and multiple assessment techniques are incorporated to create the AC. Some sort of job simulation exercise(s) should be included among these multiple techniques. Additionally, the use of multiple trained assessors is considered one of the

essential elements of an AC. Finally, all candidate behaviour should be recorded and reported in such a manner so that the data may be integrated to facilitate a selection decision on each candidate (International Task Force on Assessment Centre Guidelines, 2000).

History of the AC. The AC originated during WWII when British and German military services recruited and selected candidates with leadership potential for service as officers (More, 1987). Building on the British War Office's experiences with ACs, the US Army's Office of Strategic Services (OSS) refined their selection procedures throughout WWII, concluding with an evaluative report entitled *Assessment of Men*, which contained a number of recommendations for future AC processes (More, 1987). Intrigued and motivated by this report, Douglas W. Bray developed the first industrial application of the AC at the American Telephone and Telegraph (AT&T) Company during the 1950's (Bray, 2004). The longitudinal study, known as the Management Progress Study, investigated the AC's ability to predict managerial potential and progress throughout a person's career, the results of which helped to establish the validity of the AC process (Moses, 1977). Gradually, other organizations adopted and adapted the AT&T AC method for their own purposes (e.g., IBM, Standard Oil, General Electric, Sears; Byham, 1977). With the increased use of ACs came the need for sharing science-based research, and the development of a set of professional standards for ACs. In 1973, the first International Congress on Assessment Centre Methods (ICACM) was held to address these needs, and it has been held annually (ICACM, 2004) since that time.

Validity of ACs. The validity of any selection test or process relies upon evidence from multiple sources (e.g., existing scientific literature and information gathered from a

validation study; Catano et al., 2005). Having a greater number of evidential sources provides support for a stronger argument for the validity of a selection instrument. Unfortunately, ACs suffer from a “validity paradox” (Arthur, Woehy, & Maldegen, 2000). That is, “they demonstrate content- and criterion-related validity but lack construct-related validity” (Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004, p. 74). “Both construct and content validity are validation strategies that provide *evidence based on content*, while criterion-related validity provides *evidence based on relations to other variables*.” (Catano et al., 2005, p. 41). The ACs’ problem of low construct validity has been well established through multitrait-multimethod matrix and factor analysis (FA) research (e.g., Fleenor, 1996; Klimoski & Brickner, 1987; Lance, Lambert, Gewin, Lievens & Conway, 2004; Sackett & Harris, 1988; Schneider & Schmitt, 1992), which are two methods of examining construct validity (Crocker & Algina, 1986). In a multitrait-multimethod matrix analysis, low construct validity would manifest itself in a pattern of heterotrait-monomethod intercorrelations (i.e., discriminant validity coefficients) that are higher than the monotrait-heteromethod correlations (i.e., convergent validity coefficients; Campbell & Fiske, 1959). Therefore, low construct validity for an AC would be demonstrated by a pattern of high correlations between *different* competencies that are measured by the *same* exercise (i.e., heterotrait-monomethod intercorrelations) and low correlations between the *same* competencies that are measured by *different* exercises (i.e., monotrait-heteromethod intercorrelations). Moreover, low construct validity for an AC would be demonstrated in a factor analyses (FA) by the items clustering according to the methods used rather than by the competencies these methods were designed to measure (Campbell & Fiske, 1959).

Much AC research has found this problem of low construct validity, in that analyses of the assessor ratings indicate a common method variance rather than the measurement of different competencies (e.g., Atkins & Wood, 2002; Lance et al., 2004; Sackett & Dreher, 1982; Sackett & Harris, 1988; Schneider & Schmitt, 1992). Howard (1997) summarized the dilemma by concluding that ACs tend to measure the methods used in the ACs (e.g., tests, interviews, role plays, teamwork exercises) rather than the competencies.

Several reasons have been postulated for the lack of construct validity in ACs. Assessor bias (Hoyt, 2000), and more specifically, “halo error,” or the assessor’s tendency to give similar scores to distinct competencies (Murphy & Cleveland, 1995) may result in low construct validity. That is, the assessor may unknowingly allow the rating of one competency to influence his or her ratings on other competencies, or the assessor may be unable distinguish among the competencies (Saal, Downey, & Lahey, 1980; Turnage & Muchinsky, 1982). Some research has suggested that cognitive overload on assessors is a potential source of the error (Bycio, Alvares, & Hahn, 1987; Gaugler & Thornton III, 1989; Sackett & Dreher, 1982). In fact, Gaugler and Thornton III’s (1989) findings support the notion that people have a limited capacity to process information (Lachman, Lachman, & Butterfield, 1979). They suggested that AC creators should minimize the number of dimensions that assessors are required to process, thus reducing the potential for cognitive biases due to task complexity.

Despite the lack of construct validity found in AC research, there is support for its use as a selection method based on criterion-related evidence (Dreher & Sackett, 1981; Sackett, 1987), and, to a lesser extent on content-related validity of exercises as samples

of the job content domain (Byham, 1977; Norton, 1977). Indeed, the AC method has been found to be an effective selection and promotion tool across different types of public and private organizations, and for many different types of jobs, ranging from entry-level to executive and management positions (e.g., Campbell & Bray, 1993, Cascio & Silbey, 1979; Dayan et al., 2002; Gaugler et al., 1987; Klimoski & Brickner, 1987; Sackett, 1987). In their meta-analysis of 50 ACs, Gaugler et al. (1987) reported an overall validity coefficient of $r = .37$ in predicting job performance. However, the predictive validity coefficients for individual ACs ranged from $r = -.25$ to $r = .78$. This inconsistency in criterion-related validity results is possibly attributed to the large variability of the types and quality of procedures comprising the ACs, as well as the types and quality of criteria measures (Catano et al., 2005). Furthermore, Gaugler et al.'s (1987) meta-analysis tended to focus on ACs for managerial positions rather than on entry-level or lower-level positions, and therefore, their results may not apply to all organizations and all jobs.

Few studies have focused on the predictive validity of ACs used to select candidates for entry-level jobs. Dayan et al. (2002) suggested that ACs may be infrequently applied to entry-level positions because cognitive ability tests demonstrate an acceptable level of validity for these jobs. In their meta-analysis of the cumulative research on various predictors of job performance, Hunter and Hunter (1984) found that for entry-level jobs, there was no predictor that demonstrated validity equal to that of cognitive ability, which had a mean validity coefficient of $r = .53$.

Some researchers have investigated the incremental validity of various selection methods (i.e., personality inventories, biodata scales, structured interviews, and ACs) beyond cognitive ability, but the results have been mixed (e.g., Cortina, Doherty,

Schmitt, Kaufan, & Smith, 1992; Black, 2000; Bobko, Roth, & Potosky, 1999). With a sample of police recruits, Cortina et al. (1992) found that personality inventories, namely the Minnesota Multiphasic Personality Interview (MMPI) and Inwald Personality Interview (IPI), did not add incremental predictive validity beyond cognitive ability. However, using a similar sample, Black (2000) found that Conscientiousness, as predicted by the NEO Personality Inventory added incremental validity to cognitive testing. Also using police recruits as a sample, Pynes and Bernadin (1989) found that tests of cognitive ability had a higher correlation with training performance and job performance than did the AC ratings. Encouragingly however, Dayan et al. (2002) demonstrated that the Israeli Police Force's overall assessment rating from their AC yielded significant unique performance predictive validities beyond the test of cognitive ability. The overall assessment rating that Dayan et al. (2002) used was a "final score, given by a police psychologist, derived by clinically integrating" (p. 833) the following sources of information; (1) scores on paper-and-pencil tests (i.e., cognitive ability and personality examinations); (2) AC simulation exercises; and, (3) AC peer evaluations. However, and despite their widespread use as a selection procedure for entry-level police officers (Coulton & Field 1995), other than Pynes and Bernadin's (1989) and Dayan et al.'s (2002) research, there have been few studies examining the incremental validity of ACs beyond the variance explained by cognitive ability tests for entry-level jobs.

Cognitive Ability and Job Performance

General cognitive ability (GCA) is best understood as the ability to learn (Hunter, 1986). Specific aptitudes, such as verbal aptitude, spatial aptitude, and numerical aptitude, are cognitive abilities that are narrower than GCA, but when two or three or

more specific aptitudes are combined, they are actually a measure of GCA (Schmidt, 2002). Hunter and Schmidt (1996) argued that individual differences in the ability to learn, or GCA, play a major role in job mastery given that, for most jobs, there is a lot of information to acquire. Furthermore, Hunter and Schmidt (1996) argued that although learning is a necessary condition for performance, it is not a sufficient condition. Moreover, the ability to adapt the learning while performing the job (innovative adaptation) is also a necessary component of performance. Relying on classic learning theory, Hunter and Schmidt (1996) summarized the linkages between GCA, learning, innovative adaptation, and job performance as follows:

Because the rate and amount of learning is determined by cognitive ability, the classic theory predicts a high correlation between cognitive ability and learning. Because performance is learned, the classic theory predicts a high correlation between learning and performance. Because innovation adaptation is required by most actual work situations, the classic theory predicts that cognitive ability will be even more highly correlated with performance than would be predicted from the high correlation between ability and learning (p. 461).

Schmidt (2002) asserted that given the overwhelming research evidence, the role of GCA in training and/or job performance should not be a debate. Indeed, numerous studies have demonstrated that cognitive ability is a valid and important predictor of training and job performance (e.g., Bertua, Anderson, & Salgado, 2005; Colquitt, LePine, & Noe, 2000; Cuttler & Muchinsky, 2006; Hunter 1986; Hunter and Hunter, 1984; Ree & Earles, 1992; Ree & Earles, 1991; Ree, Earles, & Teachout, 1994; Schmidt & Hunter,

2004). It is, therefore, not surprising that cognitive ability tests are widely used in organizations for personnel selection purposes for all levels of jobs (Catano et al., 2005).

Assessment and Selection for Entry-Level Jobs

All organizations should be concerned with selecting the right candidates for entry-level positions because the costs associated with poor selection methods and decisions can be high. For example, Schmidt, Gast-Rosenberg, and Hunter (1980) estimated that the dollar differential of “good” versus “average” computer programmers was over \$10,000 per year, per programmer. This simple salary cost estimate does not account for other potential costs of poor selection decisions such as recruiting, replacement and training, loss of business, severance packages, and law suits (Catano et al., 2005; Hacker, 1996). Therefore, organizations must ensure that they utilize a reliable, valid, and legally defensible selection process, of which an AC may be a substantial component. However, ACs are costly in terms of time, money, and human resources (Catano et al., 2005).

Using the concept of utility analysis (Cascio & Silbey, 1979), a case could be made against the creation and adoption of an AC over an equally valid pre-existing selection tool. That is, if an AC explains additional variance in job performance beyond the variance explained by a cognitive ability test, it may be valuable to use as part of the selection process. If, however, an AC only provides the same predictive validity in job performance as a cognitive ability test, it is financially illogical to employ the more expensive AC. Similarly, if an AC adds only a small amount of incremental validity above and beyond a much cheaper, easier to facilitate, and less-time consuming cognitive

ability test, then the cost differential between the two selection methods should be thoroughly examined before adopting the AC.

When considering the adoption of selection methods, it is important to include the repercussions associated with making an error on the job. For many entry-level jobs, the cost of an error in job performance may be minimal in comparison to the cost associated with implementing an AC that only provides a small amount of incremental predictive validity. In the case of military, fire, police, paramedic, and other similar professions, however, the cost of poor performance may go beyond that of a quantitative value (Hunter & Schmidt, 1996). That is, in public-safety oriented organizations, the cost of a mistake or poor performance by any employee in the organization, regardless of rank level, could result in dire consequences, such as a physical or mental disability, wrongful incarceration, or death. Therefore, any incremental predictive validity, no matter how small, that can be attributed to an AC above and beyond the predictive validity of a cognitive ability test may have a significant qualitative impact upon job performance in public-safety oriented professions.

Canadian Forces Military Police Selection

In 2002, the Canadian Forces (CF) Military Police (MP) branch adopted the current form of the Military Police Assessment Centre (MPAC) as its selection method for candidates applying as an entry-level MP (Tanner & Klammer, 2005). The MPAC is the final stage of a multi-stage selection process for MP candidates. Prior to being assessed at an MPAC, candidates are screened using the following sources of information: (1) The Canadian Forces Aptitude Test (CFAT), which is a test of general

cognitive learning ability, and for which MP applicants must meet a minimum score¹; (2) Service Background – Course Reports (if the candidate is a serving member of the CF); (3) Education Transcripts; (4) Standardized Letters of Reference; (5) An Enhanced Reliability Check (a formal security and credit history evaluation); and, (6) A selection interview at either a CF Recruiting Centre (for civilian applicants) or a Base Personnel Selection Office (for military applicants wishing to transfer to the MP occupation from a different military occupation). Information from these sources is rated and then combined (see Appendix B) to produce an overall score. The overall score is then used to create a top-down selection list from which applicants may be invited to attend the MPAC, depending on the number of available positions.

The MPAC was developed after consulting several Canadian police forces to determine their best practices, a review of civilian police selection processes and selection literature was conducted, and by following the guidelines and ethical considerations for assessment centre operations found in the International Congress on Assessment Centre Methods (ICAM, 2000) model (Tanner & Klammer, 2005). The MPAC follows and meets all of the essential elements of an AC as delineated by the ICAM (2000): (1) A job analysis of the MP occupation was conducted; (2) Relevant job behaviours were classified into 12 competencies; (3) Assessment tests and tasks were designed to elicit behaviours demonstrative of the competencies; (4) Multiple assessments techniques are used to measure the 12 competencies and, as depicted in the competency/method matrix (see Appendix A), each competency is measured by at least two of the six different methods; (5) Job-related simulation exercises are included; (6)

¹ Prior to August 2005, applicants were required to meet a minimum total CFAT score of the 10th percentile. Since August 2005, MP applicants must meet a minimum total CFAT score of the 40th percentile, (Hodgson, 2005).

Multiple assessors are employed; (7) A comprehensive three-day assessor training package is delivered prior to the MPAC, which includes an extensive Frame of Reference (FOR) training section and rating behaviour using the Behaviourally Anchored Rating Scales (BARS); (8) All candidate behaviour is systematically recorded; (9) Reports are completed for each test or task; and, (10) All individual candidate scores for the 12 competencies are compiled and discussed by the assessment team, and a consensus is reached for an overall summed score (Klammer & Tanner, 2002a, 2002b, 2002c, and 2003; Tanner & Klammer, 2005). The 12 MPAC competencies, their definitions, and their BARS were developed using the results of job analyses and in consultation with subject matter experts (Klammer & Tanner, 2002a). Following each MPAC, a top-down selection list is created by using the candidates' overall score. Barring any specific areas of concern with an individual candidate (e.g., admitted drug usage in contravention to the CFPM Drug Policy, CFPM, 2002, or a score lower than 3 on any one of the Integrity competency items), offers of employment as an MP were made from the top-down selection list based on the number of available positions.

The MPAC is an expensive selection tool in terms of time, human resources, and dollars. The cost of each MPAC varies with the number of candidates and assessors and with the location in which the MPAC is held. Generally, the cost of the MPAC ranges from \$20,000 – \$125,000 with an overall annual budget of approximately \$200,000. However, no validation study has been conducted to measure the psychometric properties or the accuracy of the job performance predictions made by the MPAC. Furthermore, the MPAC's ability to predict job performance, beyond what is predicted by CFAT scores,

has not been studied. In order to examine the MPAC's predictive validity, it is necessary to examine the issue of job performance and its measurement.

CF Military Police Job Performance – The Criterion Dilemma

Organizational success is, to a certain degree, dependent upon the effectiveness with which the organization's workers execute their work responsibilities (Barnes-Farrell, 2001). Consequently, valid and reliable measures are necessary to accurately evaluate workers' performance such that appropriate human resource decisions are made to maintain or improve organizational efficiency and success. Furthermore, valid and reliable job performance measures are an essential component in the selection process because the "usefulness of selection measures is assessed by how well they predict performance" (Catano et al., 2005, p. 178). The challenge, however, is to ensure that valid and reliable measures of job performance, or criterion, are established before analyzing the extent to which the selection measure predicts the performance.

Despite job performance being one of the most important dependent variables in Industrial/Organizational (I/O) Psychology, Campbell, McCloy, Oppler, and Sager (1993) argued that "the word *performance* is misused and exploited to the extreme in society at large, and is frequently butchered beyond recognition in psychology" (p.35) such that we have no common understanding of what it really is. However, in general, we can define job performance as the behaviours that are "...relevant to accomplishing the goals of an organization" (Catano, et al., 2005, p. 168).

Measuring job performance is a complex task because different jobs will require different behaviours, and these behaviours may differ between or within organizations, even when the job titles are identical or similar. For example, a small-town police

officer's job performance domain, or set of job performance behaviours (Catano et al., 2005), will likely be quite different from that of a large city police constable's. Moreover, as Froemel (1979) stated with respect to police officers' job performance, the "complexity of the job itself, the great diversity of work situations, the intangibility of the 'work product,' the public service aspects of the job, and the largely self-supervised status of the individual police officer are but a few of the factors contributing to measurement difficulties" (p. 87). Consequently, measuring individual job performance in different jobs with the same occupational title is complex because the jobs may encompass different performance domains.

Faced with the complex task of measuring their police officers' performance, the MP Branch has had to find a way to incorporate relevant job performance behaviours into measures that are separate from the centralized CF Performance Appraisal System (CFPAS). The CFPAS assesses every member of the CF by a single set of criteria regardless of their occupation (i.e., an MP is measured against the same criteria as a military cook or a naval electronics technician), and is therefore, not MP-task specific. As such, the MP branch measures specific MP job performance on MPs ability to conduct 15 job tasks upon completion of their initial MP training course (i.e., QL3); (1) enforce regulations; (2) conduct law enforcement duties; (3) respond to emergencies; (4) effect an arrest; (5) apply defensive tactics techniques; (6) conduct searches; (7) conduct general criminal/service offence investigations; (8) conduct interview; (9) conduct duties related to evidence; (10) conduct court administration; (11) conduct unit detention activities; (12) perform community policing; (13) conduct physical security surveys; (14) execute

movement plans and orders; and (15) process refugees, stragglers, prisoners of war, and detainees.

Overview and Research Goals

A valid, reliable, legally defensible, and cost effective selection procedure is an essential element of an efficient and productive organization. An AC may meet these selection procedure standards. The Israeli police force's AC demonstrated incremental predictive validity above and beyond a test of cognitive ability (Dayan et al., 2002); however, because their predictor and criteria measures differ from those of the CF's MP branch, it is necessary to examine the CF's MPAC. Therefore, the overall purpose of this study was to examine the psychometric properties of the MPAC, including its factor structure and its incremental validity beyond the predictive validity of a cognitive ability test.

According to the MPAC, 12 competencies are assessed and rated. Therefore, the MPAC should cluster into 12 factors representing the competencies, and these 12 competency-based scales should demonstrate high internal reliability. However, research has indicated that ACs suffer from common method variance (e.g., Fleenor, 1996; Lance et al, 2004), and thus, the MPAC may cluster into six factors representative of the six MPAC methods. Therefore, the first research goal is to examine the reliability and factor structure of the MPAC to determine whether the MPAC measures 12 competencies or whether it factors into the 6 methods.

Similarly, if the MPAC accurately assesses 12 competencies (i.e., demonstrates construct validity), there should be a pattern of high convergent validity coefficients (i.e., high correlations among the *same* competencies measured by *different* methods) and low

discriminant validity coefficients (i.e., low correlations among *different* competencies measured by the *same* methods). However, if the MPAC suffers from common method variance, the discriminant validity coefficients would be higher than the convergent validity coefficients, indicating no differentiation among the 12 MPAC competencies. Therefore, the second research goal is to determine whether the MPAC has a pattern of high convergent validity coefficients and low discriminant validity coefficients or whether it has a pattern of discriminant validity coefficients higher than the convergent validity coefficients (i.e., whether or not the MPAC demonstrates construct validity).

Despite questionable construct validity, the AC tends to demonstrate criterion-related validity (Gaugler et al., 1987). Furthermore, Dayan et al.'s (2002) previous AC research for entry-level jobs showed that the AC demonstrated incremental predictive validity of job performance beyond that which can be predicted by a test of cognitive ability. However, an organization should rely upon the analytic results of their own AC to determine its efficiency, reliability, and validity, and to enhance legal defensibility of their own AC as a selection procedure. Therefore, the third research goal is to assess the incremental predictive validity of the MPAC beyond the predictive validity of a cognitive ability test.

Method

Sample

Data were obtained from 323 MP candidates who completed the MPAC. These candidates had qualified to attend the MPAC after having passed the CFAT test of cognitive ability, a medical exam, a physical fitness test, a formal security and credit history evaluation, and a structured basic selection interview. Job performance criteria

data were available for 209 of these candidates, and CFAT data were available for 203 of these 209 candidates.

Measures

Cognitive Ability. Cognitive ability was assessed using the three subscales of the Canadian Forces Ability Test (CFAT): Verbal Ability, Spatial Ability, and Problem Solving Skills (a numerical test). The total CFAT is a timed power test and is comprised of 60 multiple choice items: 15 Verbal Ability items, 15 Spatial Ability items, and 30 Problem Solving items. Correct answers are tabulated for a total score. There are an English and French version of the CFAT, and MacLennan (1995) found that the full-scale CFAT had an internal reliability of $\alpha = .87$ for the English version, and $\alpha = .83$ for the French version. However, reliability estimates for timed tests must be interpreted with some caution (Crocker & Algina, 1986). Nevertheless, the CFAT has been shown to be a valid predictor of performance at basic recruit training (Black, 1999; O'Keefe, 1998), for the MP basic qualification level course (Hodgson, 2005), and for other CF Non-Commissioned Member (NCM) occupations (e.g., Scholtz, 2004).

MPAC. The MPAC is a three-day process during which candidates were assessed on 12 competencies; (1) Integrity; (2) Analytical Thinking; (3) Decision Making; (4) Personal Impact; (5) Interpersonal Skills; (6) Tolerance; (7) Conscientiousness; (8) Performance Under Stress; (9) Teamwork; (10) Practical Intelligence; (11) Oral Communication Skills; and, (12) Written Communication Skills (see Table 1 for definitions). The 12 competencies were assessed using six different methods: (1) Group Dynamics exercise; (2) Structured Interview; (3) Role Play Exercise; (4) Background Integrity Interview; (5) Skills Test; and, (6) Fact Find Exercise. In an attempt to reduce

Table 1

Military Police Assessment Centre Competencies

Competency	Definition
1. Integrity	Degree to which the candidate adheres to the values of honesty and trustworthiness and resists temptations of an unethical or illegal nature.
2. Analytical Skills	Degree to which the candidate seeks all possible relevant information in order to assess a situation; consults widely, probes the facts carefully, and critically analyzes the issue from different perspectives. Also, degree to which the candidate can perform simple numerical calculations.
3. Decision Making	Degree to which the candidate makes rational, realistic and sound decisions. Also, degree to which candidate uses initiative when making decisions and short term plans, and knows when to refer a decision to a higher level.
4. Personal Impact	Degree to which the candidate projects a good first impression, commands attention and respect, displays an air of confidence, effectively takes charge of people/situations, and is confident in own abilities and judgment, while understanding their limitations.
5. Interpersonal Skills	Degree to which the candidate is willing and able to listen to others, uses attending skills when interacting with others (e.g., establishes eye contact, paraphrases, demonstrates interest, etc.), demonstrates sensitivity, compassion and sincerity, is tactful and diplomatic, and is able to reduce tension in potentially hostile situations.
6. Tolerance	Degree to which the candidate effectively and respectfully interacts with individuals of different backgrounds (culture, ethnicity, family status, etc.), personalities, attitudes, opinions and values
7. Conscientiousness	Degree to which the candidate meets or exceeds given standards and deadlines, persistently strives for excellence, even in difficult situations, and is efficient, thorough, hardworking and dependable.
8. Performance Under Stress	Degree to which the candidate keeps own emotions under control, restrains from negative actions when provoked or when working under stressful conditions, and effectively manages stress to prevent it from negatively impacting performance.
9. Teamwork	Degree to which the candidate collaborates with others by seeking their input, encouraging their participation, and sharing information, functions well in a team setting, and understands the importance of teamwork versus competition.
10. Practical Intelligence	Degree to which the candidate possesses good judgement and common sense, is aware of surroundings; acknowledges unusual occurrences, and remains vigilant to changes, danger or warning signs, is not overly-reliant on logic but is able to react instinctively in emergencies, and respects and follows conventional methods, but recognizes that other non-traditional methods have merit in certain situations.
11. Oral Communication Skills	Degree to which the candidate clearly, accurately and concisely expresses ideas, feelings, questions and facts verbally, in both individual and group situations.
12. Written Communication Skills	Degree to which the candidate accurately and concisely describes events and presents conclusions in a written format, organizing the material in a clear, logical manner using appropriate grammar, style and language.

Modified from Tanner & Klammer, 2005

rater bias, a different assessor scored each of the six methods so that no one assessor scored a candidate twice. Scoring was completed for each competency using Behaviourally Anchored Rating Scales (BARS) on a 5-point Likert scale, ranging from 1= no ability, to 5 = superior ability. Candidate data were collected over a period of four years at eight different MPACs.

Performance Criteria. Successful MP candidates attended basic military training, and then proceeded to the basic MP course at the Canadian Forces Military Police Academy (CFPMA) (i.e., MP Qualification level 3; QL3). Candidates were required to meet a minimum standard for 15 different job performance domain criteria in order to successfully complete the MP QL3 course. Performance criteria were written tests or practical application tests concerning rules and regulations and their enforcement that the candidates learned throughout the QL3 course. Seven of these tests were graded with a percentage score, and eight tests were graded as a pass/fail. Because all candidates passed the eight pass/fail tests, there would have been no variability in these test scores, thus these performance criteria were not included in the analysis. The seven percentage graded tests encompassed the following seven job performance domains: (1) enforce laws and regulations pertaining to the Department of National Defence (DND); (2) conduct law enforcement duties; (3) effect an arrest; (4) apply defensive tactics techniques; (5) perform duties related to evidence; (6) conduct unit detention activities; and, (7) conduct physical security surveys (see Table 2 for a brief description of the these tests).

Procedures

Candidates completed the CFAT, MPAC, and MPQL3 at separate times and locations. Due to the nature of the CF recruiting and training systems, the time

Table 2

MP QL3 Tested Performance Domains

Performance Domain	Description of Test used to Assess Performance Domain
1. Enforce Laws and Regulations Pertaining to DND	A knowledge test consisting of items pertaining to the enforcement of Federal Statutes and Regulations
2. Conduct Law Enforcement Duties	A knowledge test consisting of items pertaining to the processing of impaired drivers
3. Effect an Arrest	A knowledge test consisting of items pertaining to arrest, detainment, and post arrest procedures, as well as the articulation of Military Police powers of arrest
4. Apply Defensive Tactics Techniques	A knowledge test consisting of items pertaining to the Military Police Use of Force programme and the application of Military Police defensive tactics and self defence techniques
5. Perform Duties Related to Evidence	A knowledge test consisting of items pertaining to the processing of evidence
6. Conduct Unit Detention Activities	A knowledge test consisting of item relating to regulations pertaining to CF prisons and detention barracks
7. Conduct Physical Security Surveys	A knowledge test consisting of items pertaining to DND Security Policy.

differential between measurement administrations was not held constant. Additionally, each MPAC had enough candidates to warrant two or three assessment teams. Each team was comprised of six different assessors. The assessors were not held constant between MPACs, and there were more than 80 different assessors used throughout of the time period of this study. Assessment teams were comprised of MP Non-Commissioned Members at the rank of Sergeant and above, MP Officers, and one CF Personnel Selection Officer. All assessors were required to attend and pass a three-day assessor training package, which included FOR training (e.g., practice scoring sessions with all assessors scoring paper examples and live simulations of candidates' performances in the

different methods), a review of assessor biases, and practice assessments using the BARS. Candidates' CFAT and MPAC scores and QL3 performance data were matched and merged into one data set.

Since the inception of the MPAC, several changes and refinements to the process have been made, two of which impacted upon the current study: (1) A second Group Dynamics exercise was added; (2) Originally, two sets of Role Play scores were required (i.e., one for a Role Play Integrity exercise and one for a Role Play Fact Find exercise), whereas the most current MPAC form only requires one set of scores for the Role Play Exercise and another set of scores for the Fact Find Exercise. As a result, the following two changes were made to the data; (1) the second Group Dynamics competency scores were deleted from the analysis given that data from only 50 of the 323 cases were available; and, (2) scores from the two Role Play exercises were collapsed down into one Role Play score by taking the mean of the two scores.

Each of the six methods measured several different competencies. The 12 competency variables were computed by creating a mean score made up of the competency scores from each of the relevant methods (see Appendix A for the methods and their respective competencies). For example, candidates received scores for "Integrity" on the Structured Interview, the Background/Integrity Interview, and the Role Play. The total "Integrity" score was computed by taking the mean of the integrity scores across these three methods. In some cases, assessors had assigned candidates scores for competencies in methods that were not specifically designed to measure that particular competency. For example, 50% of the cases had a score for the Analytical Skills competency as measured by the Background/Integrity Interview, even though the

Background/Integrity Interview was not specifically designed to elicit behaviour related to Analytical Skills. In order to ensure consistency in scoring across the candidates, only the competency scores from the methods designed to assess the specific competency were included in analyses. Therefore, there were a total of 44 scores assessing the 12 competencies (see Appendix A for an overview of which methods were used to assess each of the 12 competencies).

An initial screening of the data for missing data, violations of assumptions including non-linearity, non-normality, and heteroskedasticity, and for univariate and multivariate outliers was conducted using SPSS for Windows version 11.5. No major violations of assumptions were identified. However, 18 of the 44 MPAC scores had over 10 percent missing data. The majority of these scores were from the Fact Find Exercise, whose missing data percentages ranged from 21% to 58%. This missing data was due to the fact that the scoring of the Fact Find Exercise changed over the time period of the data collection. Roth (1994) cautioned that data imputation strategies artificially increase the clarity of factor structures and asserted that a pairwise or listwise deletion strategy should be used when conducting factor analysis. Therefore, correlation analyses were conducted using pairwise deletion using the 44 scores. Because a listwise deletion strategy would have resulted in an N of 16, the PCA also was conducted using pairwise deletion but using only 26 of the scores (those scores that had less than 10% of missing data).

Results

Reliability and Factor Structure of MPAC

In order to examine the internal reliability of the MPAC, I conducted two sets of reliability analyses: In the first set, I examined the reliability of each of the 12 competencies (see Table 3). Cronbach's alpha for the 12 competency scales ranged from $\alpha = .26$ to $.69$, with the minimum item total correlations ranging from $r = -.06$ to $.55$ for each of the 12 competencies. The scale measuring Oral Communications had the highest internal consistency coefficient of $\alpha = .69$. However, the inter-item correlations for this scale ranged from $r = .25$ to $r = .41$. The lowest item-total correlations for each competency ranged from $r = -.06$ to $r = .41$. In the second set of reliability analyses,

Table 3

Internal Consistency Coefficients and Lowest Item-Total Correlations for the MPAC's 12 Competency Scales

Competency	n	Cronbach's alpha	lowest item-total correlation
1. Integrity	233	.48	.26
2. Analytical Skills	181	.55	.34
3. Decision Making	153	.50	.26
4. Personal Impact	154	.64	.28
5. Interpersonal Skills	144	.54	.23
6. Tolerance	96	.55	.30
7. Conscientiousness	83	.61	.32
8. Stress Tolerance	73	.56	.22
9. Teamwork	279	.48	.33
10. Practical Intelligence	145	.26	-.06
11. Oral Communications	153	.69	.41
12. Written Communications	242	.40	.25

I assessed the reliability of the six methods (see Table 4). Because the Skills Test was comprised of only two items, I examined the correlation between the two items, which was $r = .05$. Cronbach's alpha for the remaining five method scales ranged from $\alpha = .82$ to $\alpha = .94$. The lowest item-total correlations for each method ranged from $r = .30$ to $r = .62$. The Fact Find Exercise scale had the lowest item-total correlation of $r = .3$. However, if the "Fact Find Exercise: Written Communication Skills" item were to be deleted, the Cronbach's alpha for the Fact Find Exercise scale would increase to $\alpha = .93$ to with a minimum item-total correlation of $r = .65$.

Table 4

Internal Consistency Coefficients and Lowest Item-Total Correlations for the MPAC's Five Method Scales

Method	n	Cronbach's alpha	lowest item-total correlation
1. Group Dynamics	118	.94	.62
2. Structured Interview	224	.87	.55
3. Fact Find Exercise	107	.90	.30
4. Background/Integrity Interview	290	.82	.52
5. Role Play	156	.94	.54

To examine the MPAC's factor structure, a principal component analysis (PCA) extraction with direct oblimin rotation was performed using SPSS version 14.0 on the 26 scores from the 12 competencies and five methods with a sample size ranging from $n = 284$ to 309.² Community values tended to be high, ranging from .56 to .93. An

² Because pairwise deletion is not always an ideal method for dealing with missing data, I ran two other PCAs using a mean substitution method: (1) All 44 MPAC items were included in the analysis; (2) Only the 24 items that were used in the final PCA using the pairwise deletion method were included in the analysis. Although the loading numbers were slightly different, the results of these two analyses demonstrated the same trend of items comprising the method clustering together as factors.

examination of the scree plot clearly indicated the presence of four components labelled (1) Group Dynamics, (2) Structured Interview, (3) Background/Integrity Interview, and (3) Role Play. With the exception of two items (i.e., “Skills Test: Practical Intelligence” and “Skills Test: Written Communication Skills”), item loadings in the pattern matrix were clear and high, ranging from .61 to .90 (see Table 5), and there were no complex loadings. Because the “Skills Test: Practical Intelligence” and the “Skills Test: Written Communication Skills” items did not load onto any of the four components, they were removed from the analysis. Items clustered into the remaining four methods. These four components accounted for 67% of the variance and had low to moderate intercorrelations (see Table 6).

To examine the convergent and discriminant validity coefficients of the MPAC, I used a multitrait-multimethod matrix, a construct validation procedure recommended by Campbell and Fiske (1959). Correlations among the 12 competencies using different methods of assessment (i.e., monotrait-heteromethod coefficients) were examined for evidence of convergent validity. These coefficients ranged from $r = .12, p < .05$, to $r = .39, p < .001$. The mean convergent validity coefficients ranged from $r = .09$ to $r = .30$ for each of the 12 competency scales (see Table 7).

Correlations among the six methods across the 12 competencies (i.e., heterotrait-monomethod coefficients) were examined for evidence of discriminant validity. These coefficients ranged from $r = .19, p < .05$ to $r = .74, p < .01$, with the majority of the coefficients in the range of $r = .50$ to $r = .60, p < .01$. Only four of the 170 discriminant validity coefficients were less than $r = .33, p < .01$. The mean heterotrait-monomethod

Table 5

Pattern Matrix of PCA Four Factor Extraction Direct Oblimin Rotation

	Component			
	1	2	3	4
Group Dynamics Analytical Skills	.07	.02	-.82	-.07
Group Dynamics Decision Making	-.03	-.03	-.85	-.00
Group Dynamics Personal Impact	.02	.06	-.86	.07
Group Dynamics Interpersonal Skills	-.02	-.04	-.84	-.01
Group Dynamics Teamwork	-.00	.08	-.89	.03
Group Dynamics Oral Communications	.00	-.16	-.77	.01
Structured Interview Personal Impact	.73	-.09	-.04	.03
Structured Interview Interpersonal Skills	.81	.02	.00	.09
Structured Interview Tolerance	.79	-.05	.12	.01
Structured Interview Conscientiousness	.73	.02	.02	-.08
Structured Interview Stress Tolerance	.81	.09	.03	.03
Structured Interview Teamwork	.61	.07	-.18	-.02
Structured Interview Oral Communications	.74	-.15	-.06	.04
Background Interview Integrity	-.05	-.01	.08	.73
Background Interview Personal Impact	-.02	.01	-.04	.88
Background Interview Interpersonal Skills	.08	.08	-.03	.81
Background Interview Oral Comms.	.05	-.12	-.07	.79
Role Play Analytical skills	.07	-.83	-.02	.01
Role Play Decision Making	-.01	-.83	-.02	-.02
Role Play Personal Impact	.01	-.84	-.04	-.01
Role Play Interpersonal skills	-.02	-.84	-.01	.02
Role Play Stress Tolerance	.02	-.84	.01	.01
Role Play Practical Intelligence	-.01	-.85	-.03	-.05
Role Play Oral Communications	-.02	-.84	-.02	.07
Skills Test Practical Intelligence	.02	-.02	.05	.20
Skills Test Written Communications	-.05	-.12	-.11	.23

Note. Due to pairwise deletion, sample size ranged from n = 284 – 309.

Table 6

Correlations Among MPAC Methods

Method	1	2	3
1. Group Dynamics			
2. Structured Interview	.33**		
3. Background Integrity Interview	.16**	.30**	
4. Role Play	.23**	.23**	.20**

Note. Due to pairwise deletion, sample size ranged from $n = 301 - 307$.

** $p < .01$

Table 7

Mean and Standard Deviation of the Convergent Validity Coefficients for each of the 12 MPAC Competencies

Competency	number of convergent validity coefficients	<i>M</i>	<i>SD</i>
Integrity	3	.20	.03
Analytical Skills	3	.18	.06
Decision Making	3	.21	.82
Personal Impact	10	.23	.84
Interpersonal Skills	10	.21	.07
Tolerance	6	.20	.08
Conscientiousness	6	.20	.10
Performance Under Stress	6	.16	.10
Teamwork	1	.33	- ^a
Practical Intelligence	6	.09	.13
Oral Communications	10	.31	.04
Written Communications	1	.25	- ^a

^a number of convergent validity coefficient = 1, therefore, there is no calculated *SD*

Table 8

Mean Heterotrait-Monomethod Discriminant Validity Coefficients for each of the 6 MPAC Methods

Method	number of discriminant validity coefficients	<i>M</i>	<i>SD</i>
Group Dynamics Exercise	45	.59	.08
Structured Interview	28	.48	.09
Background/Integrity Interview	6	.56	.11
Skills Test	1	.05	^a
Fact Find Exercise	45	.53	.16
Role Play	25	.60	.11

^anumber of discriminant validity coefficient = 1, therefore, there is no calculated *SD*

discriminant validity coefficients for each of the of the six methods ranged from $r = .05$ to $r = .60$ (see Table 8).

Job Performance. Before the third research goal pertaining to the predictive and incremental validity of the MPAC could be assessed, it was necessary to examine the reliability of the performance measure. Because the CF uses a total score based on the seven QL3 percentage graded performance tests as the overall performance measure, a reliability analysis of the overall performance measure was conducted to determine the feasibility of using a total score as the criterion measure. Cronbach's alpha for the total score scale was $\alpha = .35$, with an item-total correlation ranging from $r = -.04$ to $r = .39$. Therefore, a total score could not be used as the criterion measure for subsequent analyses. The inter-item correlations were examined (see Table 9) for the presence of subscales. The low (and sometimes negative) inter-item correlations indicated that the seven QL3 performance domains were distinct. Therefore, the seven individual performance tests were used as individual outcomes.

Table 9

Correlations Among QL3 Performance Tests

	1	2	3	4	5	6
Test						
1. Enforce Laws and Regulations Pertaining to DND						
2. Conduct Law Enforcement Duties	.14					
3. Effect an Arrest	.31*	.19*				
4. Apply Defensive Tactics Techniques	.21*	-.25*	.11			
5. Perform Duties Related to Evidence	.08	.35*	.13	-.42*		
6. Conduct Unit Detention Activities	.02	-.01	.00	-.05	-.01	
7. Conduct Physical Security Surveys	.27*	.04	.28*	-.02	.11	.08

Note. Due to pairwise deletion, sample size ranged from $n = 142 - 194$.

* $p < .01$

Predictive Validity

Due to their demonstrated low internal reliabilities, scores from the 12 MPAC competency scales and the QL3 total performance scale were not used in a predictive validity analysis. Instead, correlations were conducted between the five MPAC method scales and the seven individual job performance tests (see Table 10). Only 4 of the 35 correlations were significant, ranging from $r = -.16, p < .05$ to $r = .24, p < .01$. An experiment-wise alpha was calculated at $p < .001$. Because the CF uses an overall MPAC score to select the MP candidates, I investigated the correlation between the overall MPAC score with the seven QL3 performance tests (see Table 10). Only two of the 7 correlations were significant; $r = .23, p < .01$ and $r = .18, p < .05$. An experiment-wise

alpha was calculated at $p < .007$. Correlations between the CFAT and the seven job performance tests were also conducted (see Table 10). Only 6 of the 24 correlations were significant, with r 's ranging from $r = .16, p < .05$ to $r = .23, p < .01$. An experiment-wise alpha was calculated at $p < .002$. Therefore, with no substantive correlations between cognitive ability and the performance criterion or between predictors and performance criterion, investigating the incremental predictive validity of the predictor beyond cognitive ability would have produced non-significant results.

Table 10

Pearson Correlation Coefficients of MPAC Factors/CFAT and QL3 Performance Tests

	Performance Tests ^a						
	1	2	3	4	5	6	7
<u>MPAC</u>							
<u>Method Scale</u>							
Group Dynamics	.11	-.09	.13	.02	.16	-.11	.01
Structured Interview	.04	.12	.06	.03	.11	.08	.02
Fact Find	.12	.09	.03	-.01	.08	.13	.04
Background Interview	.08	.09	.24**	.00	.13	-.06	.00
Role Play	.02	-.01	.14*	.21**	.08	-.16*	.07
Overall Score	.11	.01	.23**	.10	.18*	-.06	.05
<u>CFAT</u>							
Verbal Skills	.13	.13	.13	-.06	.15	.02	.08
Spatial Ability	.13	.05	.01	.03	-.01	-.08	.06
Problem Solving Skills	.23**	.04	.18*	.11	.08	.06	.22**
Total Score	.23**	.09	.16*	.06	.10	.02	.19**

Note. n ranged from 117 in two of the Fact Find Exercise/Test cells to 194 in five of the Structured Interview/Test cells.

^aPerformance Tests: (1) Enforce Laws and Regulations Pertaining to DND; (2) Conduct Law Enforcement Duties; (3) Effect an Arrest ; (4) Apply Defensive Tactics Techniques; (5) Perform Duties Related to Evidence; (6) Conduct Unit Detention Activities; (7) Conduct Physical Security Surveys.

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

Discussion

Little research has been directed towards the predictive validity of ACs for entry-level jobs. Given that ACs are expensive selection tools, it is important to understand their role in a reliable and valid personnel selection process. A reliable and valid personnel selection tool is even more important when dealing with public safety type jobs such as police officers. Therefore, the overall purpose of this study was to examine the psychometric properties of the MPAC, including its reliability, factor structure, convergent and discriminant validity, predictive validity, and incremental validity beyond the predictive validity of a cognitive ability test.

High internal consistency and high item intercorrelations are good indications of homogeneity, or unidimensionality, of test items (Schmitt, 1996). However, the 12 competency scales generally suffered from low internal consistency and low inter-item correlations. Arguably, the Oral Communications Skills scale demonstrated acceptable internal consistency with $\alpha = .69$. However, although alpha is a necessary indication of item homogeneity, it is not sufficient. The inter-item correlations of the scale must also be examined before determining the unidimensionality of test items (Schmitt, 1996). The relatively low inter-item correlations for the Oral Communications Skills competency scale indicated that the scale is not a reliable measure of a unidimensional construct. Conversely, the results of the reliability analysis of the method scales were much more positive. The five method scales (i.e., (1) Group Dynamics; (2) Structured Interview; (3) Fact Find Exercise; (4) Background/Integrity Interview; and, (5) Role Play) demonstrated both high internal consistency and high inter-item correlations. However, because the correlation between the two items of the Skills Test only was $r = .05$, this method scale

was unreliable. Therefore, five of the six MPAC methods demonstrated high internal reliability, which indicates that items comprising the method scales, not the competencies, were representative of unidimensional constructs.

The second part of the first research goal was to examine the MPAC's factor structure. The Fact Find Exercise was excluded from the analysis because the change in the scoring of this MPAC method resulted in a high amount of missing data. The results of the PCA indicated a clear four-component structure with the components labelled (1) Group Dynamics, (2) Structured Interview, (3) Background Integrity Interview, and (4) Role Play. The factor loadings were clear and high, and there were no complex loadings of any of the items. The two Skills Test items did not load onto any of the four components. Therefore, the results indicated that the MPAC has a four-factor structure representative of four of the MPAC methods.

The second research goal was to examine the validity coefficients of the MPAC. The results demonstrated a pattern of high heterotrait-monomethod intercorrelations (i.e., discriminant validity coefficients). These discriminant validity coefficients were also much higher than the pattern of low monotrait-heteromethod intercorrelations (i.e., convergent validity coefficients). Therefore, the multitrait-multimethod matrix results clearly indicated that the MPAC suffered a lack of both discriminant and convergent validity. The high discriminant and low convergent validity patterns suggested that the assessors were unable to distinguish between the 12 competencies while assessing behaviour within any one particular exercise. The patterns also indicate that the assessors were unable to recognize similar behaviour between exercises and appropriately categorize and rate it as a specific competency.

The MPAC was designed to measure 12 distinct competencies (i.e., constructs). Crocker and Algina (1986) suggested several methods of establishing construct validity, two of which are factor analysis (FA) and multitrait-multimethod matrix analysis. Using FA, the critical issue of demonstrating construct validity in the MPAC was to show that the items used to measure each of the 12 competencies were empirically identified as measuring a common factor. However, the results of the PCA clearly indicated a four-component structure that empirically identified items measuring four of the MPAC methods as the four components. Using the multitrait-multimethod matrix analysis to establish construct validity in the MPAC, the critical issue was to show a pattern of convergent validity coefficients (i.e., monotrait-heteromethod coefficients) higher than discriminant validity coefficients (i.e., heterotrait-monomethod coefficients). However, the multitrait-multimethod matrix of the MPAC produced a pattern of high discriminant validity coefficients that were much higher than the low convergent validity coefficients. Therefore, the results of the reliability analyses, the PCA, and the multitrait-multimethod matrix indicated that the MPAC suffered from a lack of construct validity and, instead, demonstrated common method variance.

The third research goal was to examine the predictive validity of the MPAC and to assess its incremental validity beyond the predictive validity of a cognitive ability test. Prior to assessing the predictive validity of the MPAC, it was necessary to establish the reliability of the performance measure. The overall job performance measure (based on the seven QL3 percentage graded performance tests) demonstrated low internal reliability. Additionally, the low (and sometimes negative) inter-item correlations indicated that the seven QL3 performance domains were distinct. Therefore, correlations

were conducted between the seven performance measures and the five MPAC method scales (i.e., Group Dynamics, Structured Interview, Fact Find Exercise, Background/Integrity Interview, and Role Play Exercise). Results indicated that there were four significant correlations between the predictor and the performance criteria. However, application of the more stringent and appropriate experiment-wise alpha of $p < .001$ suggested that these four correlations were not meaningful. Moreover, after applying the experiment-wise alpha of $p < .002$, the CFAT also tended to be unrelated to the seven job performance domains. The multiple regression to test the incremental validity was not conducted because with no substantive correlation between cognitive ability and the performance criteria or between predictor and performance criteria, the results would have been non-significant. However, because the CF uses an overall total score of the MPAC to select candidates for MP employment, I correlated the MPAC's overall total score with the seven performance measures and only two of these seven correlations were significant at $p < .05$. However, after applying a stringent experiment-wise alpha of $p < .007$, only one the correlations remained significant; $r = .23, p < .007$. Therefore, the overall predictive score of the MPAC also tended to be unrelated to the majority of the job performance domains of the criterion measure.

Limitations, Future Research, and Implications for the CF

The change in scoring of for the Fact Find Exercise was a limiting factor of the MPAC as was the missing data for some of the other variables. Despite pairwise deletion being a recommended method for dealing with missing data for correlation and factor analyses (Roth, 1989), it still has the propensity to attenuate the results. Consequently, the correlation coefficients may have been reduced, thus obscuring the true nature of the

relationship among competencies, methods, and factors. Accordingly, future research should ensure that all efforts are made to reduce the amount of missing data and to maintain consistency in selection methods and procedures across applicants over time.

A limitation of the MPAC process itself is the practice of encouraging assessors to rate behaviours on competencies for which the methods have not been specifically designed. In adopting this practice, some candidates may have been rewarded or punished for their verbosity and/or action/inaction while others may not. For example, if an assessor rated a candidate high on Analytical Skills following the Structured Interview because the assessor felt that the candidate stated something that was demonstrative of his or her Analytical Skills, then the candidate's average score for Analytical Skills may have increased. This extra data point would influence the overall total score of all competencies combined and therefore affect the candidate's position in the top-down selection list. Conversely, the candidate may have said something during the Structured Interview that the assessor may have felt warranted a low score on Analytical Skills, thus negatively affecting the average score for Analytical Skills and the overall total score. However, if the candidate had said nothing during that interview indicative of their Analytical Skills or if the assessor did not recognize any of the candidate's answers as demonstrative of Analytical Skill behaviour, the candidate would have been neither rewarded nor punished because the exercise was not designed to assess Analytical Skills. Furthermore, this rating procedure was not consistent across assessors. Therefore, by encouraging assessors to rate behaviours on competencies for which the methods have not been specifically designed, ratings on specific competencies may be inconsistent across candidates possibly resulting in unfair and inaccurate selection decisions.

The MPAC was designed and advanced based on sound theory and methodology (Klammer & Tanner, 2002a; Tanner & Klammer, 2005) and in accordance with the ICAM (2000) model. The 12 MPAC competencies were developed and refined using the results of job analyses and SME input, and are similar to those competencies found in some competency dictionaries (e.g., Dubois, 1993; RCMP - GRC Competency Dictionary, 2005) and police job dimension definitions (More, 1987). Furthermore, the MPAC assessor training included a detailed Frame of Reference (FOR) section, a type of training that Schleicher, Day, Mayes and Riggio (2002) found to be effective at improving the convergent and discriminant validity of assessment centre ratings. The sound competency development and use of the BARS scoring method, combined with the comprehensive assessor training demonstrates that every effort was made in an attempt to reliably measure and separate 12 different competencies. In the present study, however, the MPAC demonstrated common method variance and a lack of construct validity because; (1) the method scales had higher internal reliabilities than did the competency scales; (2) the four-factor structure was indicative of methods rather than competencies; and, (3) the high discriminant validity coefficients were higher than the relatively low convergent validity coefficients.

Common method variance refers to the portion of the correlation between two variables that results from sharing a common method of measurement (Campbell & Fiske, 1959). Other than the methods themselves, there are a number of potential reasons that the MPAC demonstrated common method variance. Obviously, high correlations across different constructs may occur when candidates exhibit the same level of performance across all competencies. Correlations within methods may also have been

inflated because there was an overlap between assessor and method (i.e., one assessor rated the various competencies across one method). Disentangling method and assessor effects was not possible because the available data set did not indicate which assessor rated which exercise. Also, because each assessor rated all the competencies for one method, halo error (i.e., the assessor's tendency to give similar scores to separate performance dimensions) may have contributed to the problem (Murphy & Cleveland, 1995). The MPAC assessor training package is a very thorough and detailed, including a comprehensive FOR portion (e.g., practice scoring sessions with all assessors scoring paper examples and live simulations of candidates' performances in the different methods) and intensive training in the use of the BARS used to score the behaviours. However, perhaps assessor training may be improved to focus on the specific problem of halo error to help reduce the impact it could have on common method variance.

Moreover, the MPAC requires assessors to measure a large number of competencies, which could contribute to the common method variance problem and thus, be a limitation of the MPAC itself. Cognitive overload on assessors could be a potential source of error (Bycio, et al., 1987; Gaugler & Thornton III, 1982; Sackett & Dreher, 1982) in that cognitive biases increase as task complexity increases. Gaugler and Thornton III (1989) found that AC assessors who were required to process a few dimensions (3) would observe and classify behaviour more accurately than assessors required to process a larger number of dimensions (6 or 9). As a solution, Gaugler and Thornton III (1989) suggested that AC creators should minimize the number of dimensions assessors are required to process, thus reducing the potential for cognitive biases. Because the MPAC assessors were required to distinguish between 12 different

competencies, cognitive overload may have been a contributing factor to the common method variance result. Therefore, future research should look at reducing the number of competencies that MPAC assessors are required to rate.

The only MPAC method that did not demonstrate common method variance was the MPAC Skills Test. The two items comprising this method did not correlate as a method scale nor did they correlate with their respective competencies. The MPAC Skills Test is comprised of two essay style questions that were designed to test candidates' ability to analyze a controversial topic, balance the argument, and provide convincing evidence, in writing, using appropriate grammar, style and language, in a clear, logical manner. In addition, the test is composed of memory and observation exercises, similar to those used in civilian police forces (Klammer & Tanner, 2002a). This test may be considered as a test designed to measure police-specific cognitive aptitudes. Research has shown that tests designed to measure specific cognitive aptitudes typically ended up being a test of general cognitive ability (e.g., Hunter, 1986; Ree & Earles, 1992, Ree et al., 1994). Therefore, if the MPAC Skills Test is supposed to be a measure of MP-specific cognitive skills, future research should address its utility and necessity as an MPAC method given that MP candidates are screened for general cognitive ability with the CFAT prior to their arrival at the MPAC.

A limiting factor of this study was the performance measure used. Crucial to a criterion-related validity study is a reliable measure of the performance criteria. Unfortunately, the overall QL3 performance measure had very low internal reliability. Moreover, the seven individual tests demonstrated no correlation with the MPAC method scales and CFAT cognitive ability test. The lack of correlation between the CFAT and the

performance measure could have been due to range restriction, in that the candidates had already been screened using a combination of scores from ratings of several sources of information (see Appendix B) including the CFAT. Consequently, the MP candidate sample could have been more homogeneous than the MP applicant population resulting in low variability among the CFAT scores, which may have clouded any correlation between the CFAT and the QL3 performance scores. Correction for range restriction on the CFAT was considered. However, selection was based on an assortment of variables for which the unrestricted variance was unknown. Because the literature does not suggest a clear range restriction solution for the scenario where unrestricted variance is unknown for one or more of the selection variables (Sackett & Yang, 2000), correction for range restriction on the CFAT was not conducted.

The generally low inter-item correlations among the seven performance tests indicated that the performance measure was more likely the problem. The negative correlations are of particular concern because the tests should be measuring positive aspects of an MP candidate's performance (i.e., the ideal MP candidate should score high on all tests). Instead, the negative correlation indicates that if an MP candidate scores high on one performance domain task, s/he scores low on another. Because the QL3 performance tests were designed to measure performance at the end of a highly cognitive-based training course and not on-the-job performance (a measure that could be more reflective of the desired MP competencies), the QL3 tests may not have been the most appropriate criterion measure for use in this study. Instead, in order to examine the predictive validity of the MPAC, future research should investigate the use of other current criterion measures (e.g., annual CFPAS Personnel Evaluation Report (PER),

quarterly CFPAS Personnel Development Reports, and MP Professional Standards Investigation Reports) or the development of a MP-specific performance measure for use at the end of the MPs' provisional employment period. Moreover, because the QL3 performance measure lacks reliability and has no meaningful correlation with the CFAT cognitive ability test, future research to investigate and improve the QL3 performance measure is necessary to accurately assess the training performance of MPs.

Additionally, the MPAC is expensive and currently costs the CF approximately \$200,000 per year to conduct (which is a conservative figure because it does not include the portion of the assessors' salaries associated with their time employed conducting the MPAC). A utility analysis of the MPAC could be compared to the results of a utility analysis of a more simple selection procedure. Cascio (1991) outlined a formula for analyzing the utility of an assessment centre. However, in order to conduct the utility analysis, the criterion-related validity of the assessment centre must be known. Only when a reliable and valid MP job performance measure is found and/or developed will a complete utility analysis (including the cost associated with a mistake or poor performance of a member of the public-safety oriented MP occupation) be possible. Therefore, future research regarding a better criterion measure is also necessary to enable the utility analysis of the MPAC.

Cochrane, Tett, and Vandercreek (2003) conducted a national survey and review of police departments within the United States regarding their police officer selection practices and procedures. Included among the reported 12 tests and procedures were; (1) interviews; (2) cognitive ability tests; (3) job simulation exercises; (4) psychological tests (e.g., the Minnesota Multiphasic Personality Inventory (MMPI), and the California

Psychological Inventory (CPI); and, (5) polygraph tests. However, Cochrane et al. (2003) emphasized that greater attention to the validation of the various selection methods within particular police departments is necessary. Furthermore, although there may be simpler and more cost effective selection tests and/or procedures, future research would be necessary in determining the validity of these tests as they pertain to the CF MP population. Moreover, the MPAC already encompasses three of the methods employed by other police departments (i.e., the interview, the cognitive ability test, and the job simulation exercise), and despite the fact that the MPAC does not demonstrate construct validity, it may demonstrate criterion-related validity using the overall MPAC score. Again, however, a reliable job performance measure is necessary for any future test validation study.

Cochrane et al. (2003) also recommended that police departments may want to specifically request the use of the CPI given its general demonstrated predictive validities. Indeed, other research (e.g., Cuttler & Muchinsky, 2003; Sarchione, Cuttler, Muchinsky, & Nelson-Gray, 1998) found that Conscientiousness, as measured by three scales of the CPI (i.e., Responsibility, Socialization, and Self-Control), significantly predicted social deviance and dysfunctional job behaviours among law enforcement officers. Because a score below “3” for integrity as measured by the MPAC results in non-selection and because the MPAC’s competency scale for integrity was not reliable, immediate attention should be given towards investigating the use of a valid and reliable measure of integrity or validating the Background/Integrity Interview against a valid and reliable measure of integrity. However, a thorough examination regarding the validity and legality of personality tests used for selection purposes is necessary before using any

one particular test. For example, in 2005 the 7th US Circuit Court of Appeals in Chicago ruled in a class action suit that the MMPI constituted a medical examination and as such violated the Americans with Disabilities Act (ADA), which limits the use of medical examinations by employers (Heller, 2005). The ADA does not apply to Canadian employers: However, because persons with disabilities are a protected group within the CF and because the CF is subject to the Canadian Human Rights Act, the use of tests such as the MMPI and other similarly derived personality tests for the purpose of screening for integrity is not recommended. Instead, future research should investigate the validity and feasibility of a personality test that would not only meet the CF requirements but would also withstand Canadian Human Rights Tribunal scrutiny (for an overview of integrity tests, see Prosser & Catano, 2006).

Aside from the MPAC, other CF occupations and other organizations have adopted or are seeking to adopt ACs as their method of selection for either entry level or specialist level employment. Immediate consideration should be given to measuring their internal reliability and to gathering multiple sources of evidence of their validity (i.e., content-, construct-, and criterion-related). Furthermore, before adopting future ACs as the selection method of choice, pilot studies should be conducted to ensure the AC demonstrates reliability and several evidential sources of validity, thus providing a sound legal basis for its use as the selection tool.

Specific Recommendations for the CF MP Branch

The primary goal of this validation study was to examine the psychometric properties of the MPAC. However, the results have revealed some areas for improvement both for the MPAC and for the performance measure used in this study. Although the

preceding section discusses the results and directions for future research in greater detail, the following summarized recommendations are specific to the CF MP Branch:

1. Investigate the use of a valid and reliable measure of integrity, and validate the MPAC Background/Integrity Interview against a valid and reliable measure of integrity.
2. Further investigate the reliability and validity of the MPAC Skills Test and if deemed necessary, examine the potential for developing a reliable MP-specific aptitude scale.
3. Examine the possibility of reducing the number of competencies that MPAC assessors are required to rate.
4. Expand assessor training on the specific problem of halo error and other rater biases.
5. To establish the criterion-related validity of the MPAC, investigate the use of other current criterion measures (e.g., annual CFPAS Personnel Evaluation Report (PER), quarterly CFPAS Personnel Development Reports, and MP Professional Standards Investigation Reports) or develop a MP-specific performance measure for use at the end of MP's provisional employment period.
6. Because of its low reliability (especially with some negative test correlations), examine the QL3 performance measure in terms of the individual items comprising each of the seven tests.

Conclusion

ACs can be reliable, valid, and legally defensible selection tools for entry-level organizational positions (Dayan et al., 2002). The MPAC is a content valid selection

procedure that was created using sound methodology. However, although the MPAC demonstrates internal reliability in five of the six methods scales, it does not demonstrate internal reliability for the competency scales. Furthermore, the PCA and MMTM results indicated that the MPAC lacks construct validity. Moreover, and most importantly, I was unable to establish the MPAC's criterion-related validity. Without strong evidence of construct and criterion-related validity, it may be difficult to support future use of the MPAC in its current form as a valid selection instrument.

References

- Arthur, W., Woehy, D.J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, 26, 813-835.
- Atkins, P.W.B., & Wood, R.E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence from 360-degree feedback programs. *Personnel Psychology*, 55, 871-904.
- Barnes-Farrell, J.L. (2001). Performance appraisal: Person perception processes and challenges. In M. London (Ed.), *How to evaluate others in organizations* (pp. 135-153). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bertua, C., Anderson, N., & Salgado, J.F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, 78, 387-409.
- Black, J. (2000). Personality testing and police selection: Utility of the "Big Five." *New Zealand Journal of Psychology*, 29, 2-9.
- Black, M.S. (1999). *The efficacy of personality and interest measures as a supplement to cognitive measures in the prediction of military training performance*. Unpublished master's thesis, Saint Mary's University, Halifax, Nova Scotia, Canada.
- Bobko, P., Roth, P.L., & Potosky, D. (1999). Derivations and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, 52, 561-589.

- Bray, D.W. (2004). Centered on assessment. Retrieved November 28, 2005, from <http://www.assessmentcenters.org/articles/centeredonassess.asp>
- Bycio, P., Alvares, K.M., & Hahn, J. (1987). Situational specificity in assessment centre ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463-474.
- Byham, W.C. (1977). Application of the assessment center method. In J.L. Moses & W.C. Byham (Eds.), *Applying the assessment center method* (pp. 31-43). Toronto, ON: Pergamon Press, Inc.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campbell, J.P., Gasser, M.B., & Oswald, F.L. (1996). The substantive nature of job performance variability. In K.R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258-299). Jossey-Bass Publishers: San Francisco.
- Campbell, J.P., McCloy, R. A., Oppler, S.H., Sager, C.E. (1993). A theory of performance. In N. Schmitt, W.C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). Jossey-Bass Publishers: San Francisco.
- Campbell, R.J., & Bray, D.W. (1993). Use of an assessment center as an aid in management selection. *Personnel Psychology*. 46, 691-699.
- Cascio, W.R. (1991). *Costing Human Resources: The Financial Impact of Behavior in Organizations*. Boston, MA: PWS-Kent Publishing Company.
- Cascio, W.F., & Silbey, V. (1979). Utility of the assessment center as a selection device. *Journal of Applied Psychology*, 64, 107-118.

- Catano, V.M., Wiesner, W.H., Hackett, R.D., & Methot, L.L. (2005). *Recruitment and Selection in Canada*. Toronto, ON: Nelson.
- CFPM Drug Policy, (2002).
- Cochrane, R.E., Tett, R.P., & Vandercreek, L. (2003). Psychological testing and the selection of police officers. *Criminal Justice and Behavior*, 30, 511-537.
- Colquitt, J.A., LePine, J.A., & Noe, R.A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707.
- Cortina, J.M., Doherty, M.L., Schmitt, N., Kaufman, G., & Smith, R.G. (1992). The “Big Five” personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology*, 45, 119-140.
- Coulton, G.F., & Feild, H.S. (1995). Using assessment centers in selecting entry-level police officers: Extravagance or justified expense? *Public Personnel Management*, 24, 223-254.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Belmont, CA: Wadsworth Group/Thomson Learning.
- Cuttler, M.J., & Muchinsky, P.M. (2006). Prediction of law enforcement training performance and dysfunctional job performance with general mental ability, personality, and life history variables. *Criminal Justice and behavior*, 33, 3-25.
- d’Arcimoles, C.-H. (1997). Human resource polices and company performance: A quantitative approach using longitudinal data. *Organizational Studies*, 18, 857-874.

- Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, 55, 827-849.
- Dreher, G.F., & Sackett, P.R. (1981). Some problems with applying content validity evidence to assessment center procedures. *Academy of Management Review*, 6, 551-560.
- Dubois, D.D. (1993). *Competency-based performance improvement: A strategy for organizational change*. Amherst, MA: HRD Press.
- Fleenor, J.W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology*, 10, 319-335.
- Foremel, E.C. (1977). Objective and subjective measures of police officer performance. In C.D. Spielberger (Ed.), *Police selection and evaluation*. (pp. 87-111). Praeger Publishers: New York, NY.
- Gatewood, R.D., & Feild, H.S. (2001). *Human Resource Selection*. Mason, OH: South-Western.
- Gaugler, B.B., Rosenthal, D.B., Thornton III, G.C., Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511.
- Gaugler, B.B., & Thornton III, G.C. (1989). Number of assessment center dimensions as a determinant of assessor generalizability of the assessment center ratings. *Journal of Applied Psychology*, 74, 611-618.
- Hacker, C.A. (1996). *The cost of bad hiring decisions and how to avoid them*. Delray Beach, FL: St. Lucie Press.
- Heller, M. (2005). Court ruling that employer's integrity test violated ADA could open door to litigation. *Workforce Management*, 84, 74-77.

- Hinrichs, J.R. (1978). An eight-year follow-up of a management assessment centre. *Journal of Applied Psychology*, 63, 596-601.
- Hodgson, K.E. (2005). *Validation of the CFAT and establishment of cutoff scores for military police selection* (Technical Note 2005-03). Ottawa, Ontario, Canada: Director Human Resources Research and Evaluation.
- Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavior and Personality*, 12(5), 13-52.
- Hoyt, W.T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64-86.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E. & Schmidt, F.L. (1996). Intelligence and job performance: Economic and Social Implications. *Psychology, Public Policy, and Law*, 2, 447-472.
- Huselid, M.A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38, 635-672
- International Congress on Assessment Center Methods (2000). *Guidelines and ethical considerations for assessment center operations*. Bridgeville, PA: Development Dimensions International.
- Johnston, P.J., & Catano, V.M. (2002). Psychomotor abilities tests as predictors of training performance. *Canadian Journal of Behavioural Science*, 34, 75-83.

- Klammer, J.D. & Tanner, L (2003). *Observations of the 2003 Military Police Assessment Centre (Jan 2003)* (Technical Note 2003-05). Ottawa, Ontario Canada: Director Human Resources Research and Evaluation.
- Klammer, J.D. & Tanner, L (2002a). *Observations of the 2002 Military Police Assessment Centre (May 2002)* (Technical Note 2002-05). Ottawa, Ontario Canada: Director Human Resources Research and Evaluation.
- Klammer, J.D. & Tanner, L (2002b). *Processing of Applicants for Military Police (MP) 811 and Military Police Officer (MPO) 81* (Technical Note 2002-07). Ottawa, Ontario Canada: Director Human Resources Research and Evaluation.
- Klammer, J.D. & Tanner, L (2002c). *The Selection of Military Police: An Assessment Centre Approach* (Conference Paper 2002-03). Ottawa, Ontario Canada: Director Human Resources Research and Evaluation.
- Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243-259.
- Koch, M.J., & Gunter-McGrath, R. (1996). Improving labor productivity: Human resource management policies do matter. *Strategic Management Journal*, 17, 335-354.
- Lachman, R., Lachman, J.L., & Butterfield, E.C., (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.
- Lance, C.E., Lambert, T.A., Gewin, A.G, Lievens, F., & Conway, J.M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, 89, 377-385.
- MacLennan, R.N. (1995). *Validity generalization across military occupational families*

(Technical Note 95-00). Toronto, Ontario Canada: Canadian Forces Applied Research Unit.

- More, H.W. (1987). Assessment and the assessment centre. In H.W. More & P.C. Unsinger (Eds.), *The police assessment center* (pp. 3-22). Springfield, IL: Charles C Thomas.
- Moses, J.L. (1977). The assessment centre method. In J.L. Moses & W.C. Byham (Eds.), *Applying the assessment center method* (pp. 3-11). Toronto, ON: Pergamon Press, Inc.
- Murphy, K.R., & Cleveland, J.N. (1995). *Understanding Performance Appraisal*. Thousand Oaks, CA: Sage Publications, Inc.
- Norton, S.D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. *Academy of Managerial Review*, 2, 442-453.
- O'Keefe, D.F. (1998). *Investigating the use of occupational personality measures in the Canadian Forces Selection System* (Technical Note 99-2). Ottawa, Ontario, Canada: Director Human Resources Research and Evaluation.
- Prosser, M.A., & Catano, V.M. *Integrity testing: Literature review and recommendations for Canadian Forces recruiting* (Literature Review 2006-01). Ottawa, Ontario, Canada: Director Human Resources Research and Evaluation.
- Pynes, J.E., & Bernardin, H.J. (1989). Predictive validity of an entry-level police officer assessment center. *Journal of Applied Psychology*, 74, 831-833.
- RCMP-GRC Competency Dictionary. (2005).
- Ree, M.J., & Earles, J.A. (1991). Predicting training success: not much more than g.

- Personnel Psychology*, 44, 321-332.
- Ree, M.J., & Earles, J.A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1, 86-89.
- Ree, M.H., Earles, J.A., & Teachout, M. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79, 518-524.
- Robertson, I.T., & Smith, M. (1989). Personnel selection methods. In M. Smith & I. Robertson (Eds.), *Advances in selection and assessment* (pp. 89-112). Toronto, ON: John Wiley & Sons.
- Roth, P.L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-560.
- Saai, F.E., Downy, R.G., & Lahey, M.A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88, 413-428.
- Sackett, P.R. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology*, 40, 13-25.
- Sackett, P.R., & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401-410.
- Sackett, P.R., & Harris, M.M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology*, 3, 214-229.
- Sackett, P.R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112-118.
- Sarchione, C.D., Cuttler, M.J., Muchinsky, P.M., & Nelson-Gray, R.O. (1998). Prediction of dysfunctional job behaviors among law enforcement officers. *Journal of Applied Psychology*, 83, 904-912.

- Schleicher, D.J., Day, D.V., Mayes, B.T., & Riggio, R.E. (2002). A new frame for frame-
of reference training: Enhancing the construct validity of assessment centers.
Journal of Applied Psychology, 87, 735-746.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why
there cannot be a debate. *Human performance, 15*, 187-210.
- Schmidt, F.L., & Hunter, J. (2004). General mental ability in the world of work:
Occupational attainment and job performance. *Journal of Personality and Social
Psychology, 86*, 162-173.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in
personal psychology: Practical and theoretical implications of 85 years of research
findings. *Psychological Bulletin, 124*, 262-274.
- Schmidt, F.L., Gast-Rosenberg, I.F., & Hunter, J.E. (1980). Validity generalization
results for computer programmers. *Journal of Applied Psychology, 65*, 643-661.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*,
350-353.
- Schneider, J.R., & Schmitt, N. (1992). An exercise design approach to understanding
assessment center dimension and exercise constructs. *Journal of Applied
Psychology, 77*, 32-41.
- Scholtz, D.C. (2004). *Validation of the CFAT and establishment of cutoff scores for
steward selection* (Technical Note 2004-01). Ottawa, Ontario, Canada: Director
Human Resources Research and Evaluation.
- Tanner, L.M., & Klammer, J.D. (2005). The “how to” guide for conducting the Military
Police assessment centre. *Unpublished Technical Note 2005*. Ottawa, Ontario,

Canada: Director Human Resources Research and Evaluation, National Defence Headquarters.

Turnage, J.J., & Muchinsky, P.M. (1982). Transsituational variability in human performance with assessment centers. *Organizational Behavior and Human Performance*, 30, 174-200.

Van Iddekinge, C.H., Raymark, P.H., Eidson, C.E., & Attenweiler, W.J. (2004). What do structured selection interviews really measure? The construct validity of behavior description Interviews. *Human Performance*, 17, 70-93.

Appendix A

MPAC Competency Measures

COMPETENCY	GROUP DYNAMICS (I)	STRUCTURED INTERVIEW	SKILLS TEST	FACT FIND EXERCISE	B/I INTERVIEW	ROLE PLAYS	GROUP DYNAMICS (II)
INTEGRITY		X			X	X	
ANALYTICAL SKILLS	X			X		X	
DECISION MAKING	X			X		X	
PERSONAL IMPACT	X	X		X	X	X	
INTERPERSONAL SKILLS	X	X		X	X	X	
TOLERANCE	X	X		X		X	
CONSCIENTIOUSNESS	X	X		X		X	
PERFORMANCE UNDER STRESS	X	X		X		X	
TEAMWORK	X	X					
PRACTICAL INTELLIGENCE	X		X	X		X	
ORAL COMM SKILLS	X	X		X	X	X	
WRITTEN COMM SKILLS	N/O	N/O	X	X	N/O	N/O	N/O

Appendix B

MPAC Pre-Selection Scoring Guide

1. Because the assessment centre process is a costly one, a pre-screening process has been devised for those individuals who wish to be considered for the MP occupation. The pre-screen will be based on those attributes considered important to be successful as an MP, but **not** scored/assessed at the MPAC.
2. The screening tool consists of the following:
 - a. Learning Potential/Ability,
 - b. Performance, and
 - c. Person-Environment Fit.
3. Each of the above areas is considered important in determining an individual's chance of success during training. The first, Learning Potential/Ability is comprised of: a) Canadian Forces Aptitude Test (CFAT) score, b) performance on military courses (for serving CF members, Regular or Reserve Force), and c) education. The second, Performance is based on Personnel Evaluation Report (PER) scores for serving members and ratings provided on all applicants' Standardized Letters of Reference (SLORs). The last, Person-Environment Fit is also composed of three parts: a) knowledge, or the realistic sense an applicant has of the type of work they are getting into; b) community service, which is any time spent volunteering within the community; and lastly c) employment history, or the amount and type of work a candidate has engaged in. Points are allotted for each overall section, and as a premise for most sections, the more police-related the applicant's efforts (whether in courses, community service, etc.), the more points they will be awarded.

Learning Potential/Ability

4. An assessment of an applicant's learning potential or ability based on an applicant's: a) CFAT results, b) results on service courses, and c) performance in an educational setting. Educational achievement is further divided into two components: a) the relatedness of their program/courses to police work, and b) the applicant's level of performance in those courses.

Note. There are two “Part C’s” – one is for Civilians and CT, the other for OTs. Please ensure you have used the correct one.

Part A: CFAT Results An assessment of the candidate’s learning potential based on their total CFAT percentile score.	SCORE	DESCRIPTIONS
	0	CFAT score below 10 th percentile.
	1	CFAT score between 10 - 27 percentile.
	2	CFAT score between 28 - 45 percentile.
	3	CFAT score between 46 - 63 percentile.
	4	CFAT score between 64 - 81 percentile.
	5	CFAT score between 82 - 99 percentile.
Part B: Service Background/Course Performance Assessment Strong Performance: top 1/3 or A grade Good Performance: mid 1/3 or B grade Average Performance: low 1/3 or C grade	SCORE	DESCRIPTIONS
	1	Consistently Average performance on military courses.
	2	Consistently Average to Good performance on military courses.
	3	Consistently Good performance on military courses.
	4	Consistently Good to Strong performance on military courses.
	5	Consistently Strong performance on military courses.
Part C - Civilian and CT applicants: Education Grade Chart Outstanding: A+ or > 90% Superior: A or 80-89% Above Average: B or 70-79% Average: C or 60-69% Below Average: D or 50-59% Unacceptable: F or < 50%	SCORE	DESCRIPTIONS
	0	Unacceptable performance in program.
	1	Below Average performance in program.
	2	Average performance in program.
	3	Above Average performance in program.
	4	Superior performance in program.
	5	Outstanding performance in program.

For W/BPSO’s: To assign a total score for Learning Potential/Ability, add together the scores awarded for Parts A, B, and C, and then divide by 3 to have a total score out of 5.

For CFRC’s: To assign a total score for Learning Potential/Ability, add together the scores awarded for Parts A and C, and then divide by 2 to have a total score out of 5. If you have a Reservist, follow directions above for W/BPSOs.

Part C - OT Applicants:	SECTION	SCORE	DESCRIPTIONS
<p>Section 1: Number of Courses Completed</p> <p>Section 2: Type of Courses Completed Preferred Subjects: Criminology, Law & Security, Police Foundations, Computer Science, Business.</p> <p>Acceptable Subjects: Psychology, Sociology, Political Science, Customs & Excise, Computer Science-related Engineering, Military Studies.</p> <p>Other Subjects: English, Engineering (other than computer science related), etc.</p> <p>Section 3: Education Grade Chart Outstanding: A+ or > 90% Superior: A or 80-89% Above Average: B or 70-79% Average: C or 60-69% Below Average: D or 50-59% Unacceptable: F or < 50%</p> <p>(Grade Point Average (GPA) equivalents are not provided because not all GPAs are based on 4.0)</p> <p>Note: This section is to be used for OT applicants who are not presently required to have a 2 yr law and security diploma.</p>	1	0	No courses completed.
		1	1 - 2 x 1/2 courses completed.
		2	3 - 4 x 1/2 courses completed.
		3	5 - 6 x 1/2 courses completed.
		4	7 - 8 x 1/2 courses completed.
		5	More than 8 x 1/2 courses completed.
	2	1	Course subjects are primarily in the Other fields.
		2	Course subjects are in the range of Other to Acceptable.
		3	Course subjects are primarily in the Acceptable fields.
		4	Course subjects are in the range of Acceptable to Preferred.
		5	Course subjects are primarily in the Preferred fields.
	3	0	Unacceptable performance on most courses.
		1	Below Average performance on most courses.
		2	Average performance on most courses.
		3	Above Average performance on most courses.
		4	Superior performance on most courses.
		5	Outstanding performance on most courses.
	To assign a total score for Educational Achievement, add together the scores awarded for Parts A, B, and C, and then divide by 3 to have a score out of 5.		

Performance

5. An assessment of an applicant's achievement in a work related context. This assessment is based on an applicant's performance as indicated on their Standardized Letters of Reference (SLOR), and/or their Personnel Evaluation Reports.

Part A: Standardized Letter of Reference Overall Assessment An assessment of an applicant's competencies, in terms of the scores received on their SLORs. Descriptions are based on an average of SLORs received.	SCORE	DESCRIPTIONS
	1	Average overall assessment of 1 on SLORs.
	2	Average overall assessment of 2 on SLORs.
	3	Average overall assessment of 3 on SLORs.
	4	Average overall assessment of 4 on SLORs.
	5	Average overall assessment of 5 on SLORs.
Part B: Personnel Evaluation Report An assessment of an applicant's work ability, in terms of the scores received on their previous PERs. Descriptions are based on an average of last 3 PERs.	SCORE	DESCRIPTIONS
	0	Scores for performance consistently in the Unacceptable range, and consistently Low potential scores.
	1	Majority of scores for performance in the Needs Improvement range. Potential scores consistently in the Low to Normal range.
	2	Majority of scores for performance in the Developing range. Potential scores consistently Normal.
	3	Majority of scores for performance in the Skilled range. Potential scores consistently in the Normal to Above Average range.
	4	Majority of scores for performance in the Exceeded Standard range. Potential scores consistently Above Average.
	5	Majority of scores for performance in the Mastered range. Potential scores consistently in the Above Average to Outstanding range.

For W/BPSO's & CFRC's: To assign a total score for Performance, indicate the score for Part A (SLOR score). The MP Branch has access to PER scores, and this component will be added to the SLOR score and divided by 2 to reflect an overall score out of 5.

Person-Environment Fit

6. An assessment of an applicant's suitability to join the MPO occupation based on their: a) knowledge of the MOC, b) employment history, and c) community service. Both employment history and community service are further divided into two components: a) the amount of time spent working or volunteering, and b) the extent to which the employment or community service is police-related.

Part A: Knowledge of MOC	SCORE	DESCRIPTIONS
	1	Individual not concerned with what occupation they receive, so long as it is not their current occupation or so long as they get a job, OR they are primarily motivated out of a need for power or status that is associated with being an MP.
	2	Below Average knowledge or completely unrealistic expectations as to the training and employment of the MP Branch, i.e., focussed on a limited aspect of MP employment such as investigative work.
	3	Average knowledge of the MP Branch, however may have unrealistic expectations as to the job, training, etc. Little or no contact with members of the branch.
	4	Above Average knowledge and fairly realistic view of the MP Branch, its training, employment, etc. Some contact with serving members.
	5	Superior knowledge and realistic view of the MP Branch, and the type of employment and training to be expected. Has spoken extensively with serving members of MP Branch.

Part B: Employment History An assessment of an applicant's work experience, in terms of the amount of previous experience an applicant has, and the relatedness of the applicant's previous work experience to police employment. Definitions PT: part-time FT: full-time 1 year PT = 6 months FT; 2 years PT = 1 year FT, etc. Police-Related: Employment in Federal, Provincial or Municipal law enforcement or security organizations (e.g., RCMP, OPP, Ottawa Regional Police, Corrections, CSIS, CCRA, etc.) Semi-Police Related: Employment as a Security Guard, Commissionaire, Private Investigator, etc.	SCORE	DESCRIPTIONS
	0	Less than 6 months FT experience in non-police related work.
	1	6 months to 1 year FT experience in non-police related employment, OR less than 6 months FT experience in semi-police related work.
	2	1 to 2 years FT experience in non-police related work, OR 6 months to 1 year FT experience in semi-police related work, OR less than 6 months FT experience in police-related work.
	3	2 to 3 years FT experience in non-police related work, OR 1 to 2 years FT experience in semi-police-related work, OR 6 months to 1 year FT experience in police related work.
	4	3 or more years FT experience in non-police related work, OR 2 to 3 years FT experience in semi-police related work, OR 1 to 2 years FT experience in police-related work.
	5	3 or more years FT experience in semi-police related work, OR 2 to 3 years FT experience in police related work.

Part C: Community Service Relevant experience includes such things as: victims services, victim assistance programs, police ride-along, outreach programs (e.g., food shelters, homeless shelters, eldercare), working with teens/Girl Guides/Scouts/Cadets/Big Brother/Big Sister, help lines, suicide prevention programs, hospitals, etc. Participation Regular: 1 to 2 times per month Occasional: 1 to 2 times per year	SCORE	DESCRIPTIONS
	0	No community service or volunteer work.
	1	Less than 1 year occasional participation in a non-police related organization.
	2	Greater than 1 year occasional participation in a non-police related organization, OR less than 1 year occasional participation in a police related organization.
	3	Less than 1 year regular participation in a non-police related organization, OR greater than 1 year occasional participation in a police-related organization.
	4	Greater than 1 year regular participation in a non-police related organization, OR less than 1 year regular participation in a police related organization.
	5	Greater than 1 year regular participation in a police related organization.

MPAC PRE-SELECTION SCREENING FORM**APPLICANT SN:** _____**DATE:** _____**APPLICANT SURNAME & INITIALS:** _____

		SCORE	AVERAGE
Learning Potential/Ability	Part A: CFAT		
	Part B: Service Background		
	Part C: Education		
Performance	Part A: SLORs		
	Part B: PERs		
Person-Environment Fit	Part A: MOC Knowledge		
	Part B: Employment History		
	Part C: Community Service		
TOTAL = Sum of the Average Scores			

Modified from Klammer & Tanner, 2002b



**Saint Mary's
University**

Halifax, Nova Scotia
Canada B3H 3C3

Patrick Power Library

tel 902.420.5534

fax 902.420.5561

web www.stmarys.ca

Research Ethics Board Certificate Notice

The Saint Mary's University Research Ethics Board has issued an REB certificate related to this thesis. The certificate number is: 05-110

A copy of the certificate is on file at:

Saint Mary's University, Archives
Patrick Power Library
Halifax, NS
B3H 3C3

Email: archives@smu.ca

Phone: 902-420-5508

Fax: 902-420-5561

For more information on the issuing of REB certificates, you can contact the Research Ethics Board at 902-420-5728/ ethics@smu.ca .

w h e r e t r a d i t i o n m e e t s t h e f u t u r e