The Validity of the California Psychological Inventory in the Prediction of Police Officer Applicants Suitability for Employment

Casey O. Stewart

Pacific University

Recommended Citation

This Dissertation is brought to you for free and open access by the College of Health Professions at CommonKnowledge. It has been accepted for inclusion in School of Graduate Psychology by an authorized administrator of CommonKnowledge. For more information, please contact CommonKnowledge@pacificu.edu.
The Validity of the California Psychological Inventory in the Prediction of Police Officer Applicants Suitability for Employment

Abstract
This study explores the relationship between the California Psychological Inventory Suitability Risk Levels (Roberts and Johnson, 2001) and police applicant suitability as determined by background investigators during the preemployment selection process. The CPI Suitability Risk Levels are numerical probabilities indicating the likelihood that an individual will be rated “poorly suited” by expert screening psychologists, terminated after being hired, or found to have engaged in a variety of problem behaviors (e.g., integrity violations, illegal drug use, criminal behavior, poor work history, etc.). Suitability determinations were made based on ten non-medical objective dimensions related to the essential job functions of law enforcement officer. The results indicate that two of the eight CPI Risk Levels, Probability of involuntary departure and Poorly suited, were moderately related to background investigators’ suitability determinations. Further, the CPI risk estimate Probability of involuntary departure was the best predictor of police applicant suitability determination. However, the prediction model failed to reach statistical significance. Nevertheless, classification analyses revealed that the CPI risk estimates did a good job in correctly predicting suitability judgments on the basis of background investigations. The CPI risk estimates correctly classified 74 percent of cases. All other CPI risk estimates were weakly related to, and adequate to poor predictors of, background investigation suitability determinations. Discussion of the practical application and economic utility of the CPI in screening police applicants has relevance outside of this data sample, as all police agencies struggle with selection issues. This project supports the use of personality measures in selecting suitable police applicants.

Degree Type
Dissertation

Degree Name
Doctor of Psychology (PsyD)

Committee Chair
Jay C. Thomas, Ph.D., ABPP, Chair

Second Advisor
David M. Corey, Ph.D., ABPP

Third Advisor
Michel Hersen, Ph.D., ABPP

Subject Categories
Psychiatry and Psychology

This dissertation is available at CommonKnowledge: https://commons.pacificu.edu/spp/156
Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu
THE VALIDITY OF THE CALIFORNIA PSYCHOLOGICAL INVENTORY IN THE
PREDICTION OF POLICE OFFICER APPLICANTS SUITABILITY FOR
EMPLOYMENT

A DISSERTATION
SUBMITTED TO THE FACULTY
OF
SCHOOL OF PROFESSIONAL PSYCHOLOGY
PACIFIC UNIVERSITY
HILLSBORO, OREGON

BY
CASEY O. STEWART

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY
JULY 25TH, 2008

Jay C. Thomas, Ph.D., ABPP, Chair
David M. Corey, Ph.D., ABPP
Michel Hersen, Ph.D., ABPP
ABSTRACT

This study explores the relationship between the California Psychological Inventory Suitability Risk Levels (Roberts and Johnson, 2001) and police applicant suitability as determined by background investigators during the preemployment selection process. The CPI Suitability Risk Levels are numerical probabilities indicating the likelihood that an individual will be rated “poorly suited” by expert screening psychologists, terminated after being hired, or found to have engaged in a variety of problem behaviors (e.g., integrity violations, illegal drug use, criminal behavior, poor work history, etc.). Suitability determinations were made based on ten non-medical objective dimensions related to the essential job functions of law enforcement officer. The results indicate that two of the eight CPI Risk Levels, Probability of involuntary departure and Poorly suited, were moderately related to background investigators’ suitability determinations. Further, the CPI risk estimate Probability of involuntary departure was the best predictor of police applicant suitability determination. However, the prediction model failed to reach statistical significance. Nevertheless, classification analyses revealed that the CPI risk estimates did a good job in correctly predicting suitability judgments on the basis of background investigations. The CPI risk estimates correctly classified 74 percent of cases. All other CPI risk estimates were weakly related to, and adequate to poor predictors of, background investigation suitability determinations. Discussion of the practical application and economic utility of the CPI in screening police applicants has relevance outside of this data sample, as all police agencies struggle with selection issues. This project supports the use of personality measures in selecting suitable police applicants.
ACKNOWLEDGEMENTS

I would like to first thank my mother, whose constant expression of love and untiring commitment to pushing me to be the best person that I can be has served as a model of courage and morality. Mom, you are a strong woman and a beautiful person: I love you and I am deeply grateful. Next, I must thank my father, who toiled at his trade to provide for our family and made it possible for me to obtain an education and have freedom of choice for my future. I must also thank my grandparents and great grandmother, who have invested significant resources in cultivating the man I am today. They have always been present and supportive, teaching me about the world and how to be a positive part of it. I would also like to thank Jay Thomas, my dissertation chair, for his guidance and prudent consult. Jay reaffirmed for me what I learned from reading William James as an undergraduate: That critical analysis is essential for understanding, but that thinking is only as valuable as its impact in practice. One person in particular who has influenced my life in many ways is Dave Corey. Dave, you reminded me that generosity is a virtue and demonstrated how to be an ethical man in practice not just in mind: You are a fine man and an excellent friend. The most influential person in my life over the past decade is most certainly my wife, whose commitment to relationships and helping others make her the most beautiful person I know. Not only has she given needed support for my success in graduate school, but she has continued to challenge me to do more for people. My life is exponentially richer and I am sure that I am a better person for knowing her. Finally, this research would not have been possible without the generous cooperation of Mike Roberts and Mike Johnson as well as the Chief of Police and his staff from the represented department.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>13</td>
</tr>
<tr>
<td>A Brief History of Psychological Testing and Police Selection</td>
<td>13</td>
</tr>
<tr>
<td>Preemployment Psychological Evaluation of Police Candidates</td>
<td>24</td>
</tr>
<tr>
<td>Legal Issues</td>
<td>26</td>
</tr>
<tr>
<td>Assessment Procedures</td>
<td>33</td>
</tr>
<tr>
<td>Finding the Best Approach for Hitting a Moving Target</td>
<td>39</td>
</tr>
<tr>
<td>An Overview of the California Psychological Inventory</td>
<td>42</td>
</tr>
<tr>
<td>Description and Purpose</td>
<td>42</td>
</tr>
<tr>
<td>History and Development</td>
<td>44</td>
</tr>
<tr>
<td>Scales and Vectors</td>
<td>47</td>
</tr>
<tr>
<td>Peer Review and Empirical Support</td>
<td>51</td>
</tr>
<tr>
<td>Administration, Scoring, and Interpretation</td>
<td>55</td>
</tr>
<tr>
<td>Special Applications</td>
<td>58</td>
</tr>
<tr>
<td>The California Psychological Inventory and Police Selection</td>
<td>65</td>
</tr>
<tr>
<td>Study Descriptions and Findings</td>
<td>67</td>
</tr>
<tr>
<td>Limitation</td>
<td>99</td>
</tr>
<tr>
<td>Summary of the Results</td>
<td>102</td>
</tr>
<tr>
<td>THE CURRENT STUDY</td>
<td>106</td>
</tr>
<tr>
<td>Issues this Study is Tackling and Hypotheses</td>
<td>107</td>
</tr>
<tr>
<td>METHOD</td>
<td>111</td>
</tr>
<tr>
<td>Subjects</td>
<td>111</td>
</tr>
<tr>
<td>Measures</td>
<td>112</td>
</tr>
<tr>
<td>Procedures</td>
<td>113</td>
</tr>
<tr>
<td>RESULTS</td>
<td>115</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>115</td>
</tr>
<tr>
<td>Correlation Analyses</td>
<td>120</td>
</tr>
<tr>
<td>Logistic Regression Analyses</td>
<td>121</td>
</tr>
<tr>
<td>Classification Analyses</td>
<td>131</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>141</td>
</tr>
<tr>
<td>Summary of Results</td>
<td>141</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Means, Trimmed Means, Medians, Standard Deviations, and Percentiles of CPI Risk Estimates</td>
<td>116</td>
</tr>
<tr>
<td>Table 2</td>
<td>Correlations Among Predictor Variables</td>
<td>121</td>
</tr>
<tr>
<td>Table 3</td>
<td>Logistic Regression Analysis of Background Investigation Outcome on CPI Variables (After Controlling for Ethnicity, Age, and Sex)</td>
<td>124</td>
</tr>
<tr>
<td>Table 4</td>
<td>Logistic Regression Analysis of Background Investigation Outcome on CPI Variables (Without Controlling for Ethnicity, Age, and Sex)</td>
<td>125</td>
</tr>
<tr>
<td>Table 5</td>
<td>Hierarchical Logistic Regression Analysis of Background Investigation Outcome on CPI variables (After Controlling for Ethnicity, Age, and Sex)</td>
<td>128</td>
</tr>
<tr>
<td>Table 6</td>
<td>Classification Table for Hierarchical Logistic Regression Analysis for Background Outcomes (After Controlling for Ethnicity, Age, and Sex)</td>
<td>133</td>
</tr>
<tr>
<td>Table 7</td>
<td>Classification Table for Hierarchical Logistic Regression Analysis for Background Outcomes (Without Controlling for Ethnicity, Age, and Sex)</td>
<td>135</td>
</tr>
<tr>
<td>Table 8</td>
<td>Classification Values for Single Predictor Logistic Regression Analyses for Background Outcomes (After Controlling for Ethnicity, Age, and Sex)</td>
<td>136</td>
</tr>
<tr>
<td>Table 9</td>
<td>Classification Values for Single Predictor Logistic Regression Analyses for Background Outcomes (Without Controlling for Ethnicity, Age, and Sex)</td>
<td>138</td>
</tr>
</tbody>
</table>
INTRODUCTION

Selecting the most suitable applicants for the job of law enforcement officer is a costly endeavor. Personnel costs consume about 85 percent of the local law enforcement agency budget (Bradford, 1998). The New York Police Department estimates that each new officer costs about $500,000 (Decicco, 2000). According to Cochrane et al. (2003), it costs a large metropolitan police department approximately $100,000 to train each new police recruit. Further, Fitzsimmons (1986) reported that it costs a major city almost a half million dollars for each hiring error that results in an unsuitable officer.

Millions of dollars have been lost through litigation because of discriminatory hiring practices (Horstman, 1976). Beyond the monetary loss and waste of human resources that accompany an erroneous hiring decision and, more importantly, is the significant liability to both the public safety and the integrity of the hiring agency of selecting an applicant that is unqualified for the job of law enforcement officer. Additionally, not selecting a suitable individual for reasons unrelated to the essential functions of the job (e.g., age, sex, ethnicity, etc.), intentionally or not, reflects a major social injustice and is illegal (e.g., employment discrimination).

The financial investment in getting a probationary officer on the street as well as the financial liability of selecting the wrong individual for the job highlights the importance of the selection process. Psychological screening alone costs more than $150 per police officer applicant (Ash, Slora, & Britton, 1990) and is a standard practice in most law enforcement agencies (Varela, Boccaccini, Scogin, Stump, and Caputo).

Bartol and Bartol (2004) reported that over 18,000 organizations in the United States qualify as law enforcement agencies. These agencies exist at the federal, state,
county, and local or municipal level. Law enforcement agencies are paramilitary organizations and the job of police officer is substantially different than most occupations in the private sector (Super, Blau, Wells & Murdock, 1993), qualifying “… as one of the most complex in our society” (Baehr et al., 1968, p. 226). Law enforcement officers are entrusted with the great responsibility of maintaining social order and protecting the public from harm. They are the “gatekeepers of the criminal justice system” (Scrivner, 1994). Officers frequently carry out their duties in extremely demanding and stressful conditions where they themselves are placed in harms way. Police officers are also authorized to use force if warranted, which can result in brutal and fatal consequences (Skolnick, 2000). Bittner (1970) noted that police officers are the only agents of society licensed to use lethal force prior to adjudication. Carlson (1975) noted that police have the authority to use exercise power that “may literally save or destroy individuals” (p. 2).

In addition to the stressful and potentially dangerous nature of the job, law enforcement work is often done in an atmosphere where public opinion of police officers is low (Varela, Boccaccini, Scogin, Stump & Caputo, 2004). Few occupations have been the object of such public attention, controversy, and debate as that of law enforcement (Baehr et al., 1968). Headlines and media reports of police misbehavior and abuse of power are a major concern of law enforcement agencies and have a significant impact on the public trust and, thus, officers’ ability to effectively police.

The complex relationship between agents of the law and the citizenry makes sense considering the dynamics between those charged with enforcing the rule of law and those required to abide. In an autobiography about his career working for the New York Police Department, Edward Conlon (2004) wrote “there were those who saw us as their
protectors, and those who saw us as their keepers, and both were right” (p. 11). For those individuals in the latter category, such a complex relationship can presuppose some degree of volatility. The tentative relationship between keeper and the kept can be made less stable by a number of circumstances. One factor that contributes to the destabilization of relations between law enforcement and the citizenry is incidents involving the excessive use of force (Bartol & Bartol, 2006). Such incidents not only reduce the faith of citizens in law enforcement personnel, but create fear in the community. Another factor that contributes significantly to such unstable relations is the problem of corruption among police officers (Arrigo & Claussen, 2003). According to (McCafferty, Souryal & McCafferty, 1998) corruption in law enforcement has been pervasive and continues to be a serious problem in many police departments. Corruption not only negatively impacts the public trust, but such misconduct harms other law enforcement personnel, the police institution itself, and stakeholders.

Based on the heightened potential for manifold negative consequences, the job of police officer is considered a high-risk occupation (Borum, Super, & Rand, 2003). The sensitive nature of this position has caused law makers, administrators, and professionals alike to develop detailed guidelines for the preemployment assessment of police officer applicants. Identifying and selecting competent police officers has been a critically important social issue addressed by experts from varying fields of scientific inquiry. Social scientists interested in assessment and personality have been particularly active in this scientific pursuit (Mills & Bohannon, 1980). A standard practice among law enforcement agencies is to employ some type of psychological assessment instruments to aid in the selection of job candidates (Borum & Stock, 1993).
Psychologists have been conducting preemployment psychological evaluations since the early 1900s. In recent years, psychologists have been increasingly active in conducting these types of evaluations for law enforcement agencies (Borum et al., 2003). The courts have ruled that law enforcement agencies not only have the right to conduct psychological evaluations, but may be held liable for the actions of officers who were not properly screened (Super, 1999). Therefore, most major law enforcement agencies employ psychological testing as one element of their multistage selection systems (Varela, et al., 2004).

The purpose of preemployment psychological screening is to provide relevant information to hiring agencies about those candidates who may be more likely to engage in counterproductive work behavior and those who may pose a significant public safety risk (Janik, 1994). Hargrave and Hiatt (1989) observed that beginning sometime around the 70s, a major focus of the police selection movement was the use of personality measures to assess characteristics related to job performance. Personality testing is one mechanism for identifying officer candidates that may be unable to perform their job duties under the typical work conditions (Varela et al., 2004). One personality measure that has recently been made available with special norms for various classes of public safety applicants and incumbent employees (e.g., police, fire/EMS, corrections, and emergency communications dispatchers) is the California Psychological Inventory (Gough & Bradley, 1996).

According to Hargrave and Hiatt (1989), the CPI is one measure that appears to be particularly suited for police selection because of its ability to evaluate normal personality variables important for social interacting. The CPI is considered important to
police selection because police work involves regular contact with people in a multitude of situations making the assessment of interpersonal skills critical. The CPI is one of the most frequently utilized assessment instruments for evaluating the psychological suitability of police officer applicants (Cochrane, Tett & Vendecreek, 2003), second only to the MMPI-2, which is the most widely used instrument for psychological assessment (Borum, et al., 2003; Cochrane et al., 2003). The CPI consists of 18 primary scales that are anchored in normal-range behavior. According to Blau (1994) the scale constructs have both a face valid and empirical relationship to the quality of law enforcement officers’ job performance (e.g., Tolerance, Responsibility, Empathy, Self-Control, Flexibility etc.). The empirical literature related to the validity of the CPI for use in a police and public safety context is hampered by the single most persistent limitation facing predictive validity research: namely, selection bias, which occurs when the predictor under study is used to select the sample under study. Thus, selection bias serves to distort the significance of any true relationship between CPI data and police applicant or police officer characteristics whenever the subject pool is refined or reduced by use of CPI data (Corey, personal communication, 2007; Ben-Porath, 2003).

The present study is an examination of the validity of the CPI when applied to the prediction of police applicant qualification as determined by background investigators using a blind procedure in which no CPI data were used when forming judgments about applicant qualification. To the extent that the CPI is shown to have predictive validity in this context, this evidence may support the use of the CPI as a selection tool very early in the evaluation of police officer applicants, an outcome with potential widespread benefits. One example of the value of this study is the large financial savings that results
from screening out unsuitable applicants earlier in the selection process; savings that can instead be devoted to other public safety endeavors.
BACKGROUND

A Brief History of Psychological Testing and Police Selection

There are hundreds of different well refined and sophisticated psychological tests that have been developed to measure constructs such as personality and cognitive abilities. According to Ainsworth (1995), psychological tests can be used in personnel selection to determine which applicants are suitable and which applicants will prove problematic for purposes of employment.

The use of psychological testing in the prediction of future performance has been documented as early as 1300 B.C.E. The Israeli army operating under Gideon utilized military aptitude tests to select suitable soldiers. The Greek army, as observed by Plato, also implemented military ability testing to screen potential soldiers (Guion, 1976). More recently, some of the earliest published psychological tests were measures of mental ability, which were developed by Cattell (1890) for examining which mental abilities were related to college success. Blau (1994) noted that psychological testing was used during World War I for selecting military personnel. The Army Alpha and Beta tests were developed to classify military recruits. The Army Alpha and Beta tests were published in January of 1919, and by the end of the war they had been administered to approximately two million men (Larson, 1994; McGuire, 1994). These measures became the model for test development resulting in instruments like the Army General Classification Test (AGCT), which was designed and implemented for use during World War II. Even before the United States involvement in WW I, Thorndike reported on instruments for predicting job performance, specifically the performance of salespeople.
As early as 1916 the Army Alpha test was used as a screening instrument for the Pennsylvania State Police (Yoder, 1942). The use of psychological tests for predicting performance has increased and expanded since 1945. Tests have been developed to measure educational capacities, classify military and civil service personnel, and some tests have been specifically developed for private industry (Elam, 1983). Humm and Humm (1950) reported on early attempts to use personality measures for predicting police officer performance. After the 50s, the use of psychological testing in police officer selection developed into a major area of research interest (Blau, 1994). Hargrave and Hiatt (1989) observed that beginning sometime around the 70s, law enforcement selection became a major social issue and that a main focus of the police selection movement was the use of personality measures to assess characteristics related to job performance. Pugh (1985) noted that personality factors are frequently cited as important in the selection of police officer applicants. The use of personality testing for screening became widespread in the 1970s (McCreedy, 1974), although it was not until the 1980s that police departments across the country began to utilize psychological testing and interviewing (Blau, 1994).

Cronbach (1949) noted that an attempt to predict underlies every use of testing. According to Ainsworth (1995), “Psychological tests are useful in deciding whether or not a person possesses certain qualities or attributes” (p. 140). The most important reason for the use of psychological testing or evaluation in a selection process is to predict how individuals will perform on the job (Knights, 1976). According to Hibler and Kurke (1995), “The validity of a selection instrument or of any selection decision-making
Scores on personality tests are thought to relate to an individual’s future performance (Ainsworth, 1995). Most formal definitions of personality refer to the distinct constellation of character traits that define an individual and determine that person’s pattern of interaction with the environment (Gowan & Gatewood, 1995; Allport, 1961). If one accepts this definition, then it makes sense to consider the use of personality as a predictor of future performance (Elam, 1983), particularly for the job of police officer (Hogan, Carpenter, Briggs, & Hansson, 1985). The earliest personality tests were used for occupational selection and prediction (Cronbach, 1949). Although most traditional personality inventories were not developed as occupational screening instruments or job performance predictors, they are commonly used for these purposes (Elam, 1983).

Psychologists have been working in some capacity or another within law enforcement for more than 50 years in the United States (Bartol, 2006) and over 80 years in other countries (Viteles, 1929). Bartol and Bartol (2004) reported that there have been four distinct trends identifiable in the history of police psychology: (1) mental abilities testing of law enforcement officers, (2) personality assessment of police officers and the search for a “police personality,” (3) clinical services aimed at stress management, and (4) fairness in testing. One of the many services that psychologists provide to law enforcement agencies is psychological testing for purposes of preemployment screening. Blau (1994) reported that psychological testing of police officer candidates is the lion’s share of psychologists’ work in law enforcement. According to an unpublished nation-
wide survey of police psychologists conducted by Bartol in the spring of 1994, respondents indicated that the largest percentage (34.3) of their time was dedicated to preemployment screening (Bartol, 2006). Blau reported that “the psychologist’s role in recruit selection has become that of test administrator, interpreter, and interviewer” (p. 70).

The selection of police officer applicants and the involvement of psychologists in the selection process have a long history. Prior to psychologist’s involvement in the selection of individuals applying for law enforcement positions, the process was rather unsophisticated. According to Gowan and Gatewood (1995), the selection process for the Metropolitan Police in early 19th century England was very simple. In order to apply for the position of police officer an individual needed to submit a petition and two letters of recommendation to the commissioner. If the application materials were deemed acceptable, the individual was placed on an eligibility list to be considered upon the next vacancy (Grant, cited in Tobias, 1972).

At present, most major law enforcement agencies employ psychologists in the selection of police officer candidates (Varela, Boccaccini, Stogin, Stump, & Caputo, 2004). However, until relatively recently, psychologists’ involvement in police screening was the exception rather than the rule (Bartol, 2006). Psychologists have traditionally been utilized “as needed” and have not played a systematic role in personnel development and management (Scrivner, 1994). Prior to 1972, when amendments to the Civil Rights Act of 1964 extended its regulations to state and local governments, law enforcement selection was far less formalized and lacking in empirical support (e.g.,
validation). Before this time, the majority of testing was conducted using intelligence measures that assessed verbal abilities.

Blau (1994) reported that the first psychologist to engage in psychological testing of public safety applicants for purposes of selection was Lewis Terman in 1916. At that time, Terman (1917) was administering the Stanford-Binet to assess cognitive functioning in police and fire department applicants for the city of San Jose, California (Murphy, 1972). After Terman, Louis Thurston (1922) continued in the testing of police officers’ intelligence by administering the Army Alpha, a test of intelligence developed by the United States government for determining the placement of military recruits, to incumbent officers in the Detroit Police Department (Bartol & Bartol, 2004).

Policing is a complex task and a certain level of intelligence is considered to be a prerequisite for success. However, the research has shown that IQ is much less effective in predicting success in the field than academy performance (Taylor & Pease, 1988). Ainsworth (1995) claimed that although a certain level of intelligence is necessary, the most intelligent person is not necessarily the best person for the job and that other qualities may be as important if not more important. Interestingly, a Federal court dismissed a lawsuit by a police officer applicant who was rejected from an agency after being determined “too smart” for the job (Jordan v. City of New London, 1999). Although rejecting an applicant based on high scores on an intelligence test is not recommended, such practices are not a violation of federally-protected rights (http://www.aele.org/law/Digests).

Overall, cognitive testing has been found useful in predicting police academy performance but has not been particularly helpful in predicting on-the-job performance

Despite the evidence against using mental abilities tests, the cognitive approach was virtually unchallenged until the mid-60s, when the focus of testing shifted to the assessment of personality. The impetus for the use of personality testing in the selection of law enforcement officers can be seen as a confluence of multiple influences. Bartol (2006) attributed the shift from cognitive testing to personality testing to a combination of factors. The first factor affecting the shift from cognitive testing to personality testing in police selection involved concerns about the adverse impact of intelligence testing on minority groups. The second factor involved in the usage of personality assessment over cognitive testing was the Presidential Commission in Law Enforcement and the Administration of Justice (1967), which recommended that law enforcement agencies improve their selection procedures through the use of psychological tests in assessing aspects of police officer candidate’s character: Specifically, emotional stability and prejudice. According to Carlson (1975), “The US National Advisory Commission on Civil Disorder (1968) reported to the president on the role of police in escalating racial turmoil, recommending the use of psychological testing to screen out undesirable candidate for police work” (p. 18). The President’s Crime Commission of 1967 and the
President’s Riot Commission of 1968 recommended that law enforcement agencies improve the screening of police officer applicants to eliminate hiring individuals with characteristics unsuitable for police work.

In response to this recommendation, Congress devoted funds by way of the Law Enforcement Assistance Administration (LEAA). LEAA was developed in 1965 to encourage police departments to adopt more human sensitive practices and for mental health professionals to be retained in order to assist in screening out candidates who were unfit for police service due to emotional instability and/or prejudice. As a result of these events, the use of personality assessment in the screening of police officer applicants increased. However, those agencies employing psychological testing were still in the minority.

In the mid-50s, there were 30 cities with populations over 25,000 that used some type of psychiatric or psychological examination (Bartol, 2006). By the early-60s, 49 cities with similar populations were doing the same (O’Connor, 1962). The techniques being used around that time ranged from superficial paper and pencil tests to extensive test batteries and interviewing strategies. A 1963 survey by Narrol and Levitt indicated that 16% of cities surveyed used some type of psychiatric interview. However, at that time, cultural, motivational, emotional, and personality dimensions were ignored for the most part (Mills, McDevitt, & Tonkin, 1965). In another more recent study, Murphy (1972) surveyed both local and state law enforcement agencies and found that 43.9% of local and 13% of state agencies used psychological tests. Approximately 50% of those agencies used the MMPI. By the mid-70s, almost half of the larger metropolitan law enforcement agencies utilized psychological testing of some type (Murphy, 1972).
However, there appeared to be no significant change from previous studies of major police departments within the United States (Narrol & Levitt, 1963). Nevertheless, as indicated above, the preference was now for personality testing.

Bartol (2006) reported that the trend in personality assessment split into two lines of research. The first line of research was devoted to the discovery of a consistent personality profile of those individuals who select law enforcement as a career. Check and Klien (1977) conducted a literature review on the topic of police personality and found no evidence to support such a notion. The research that does exist on the topic is mostly nonscientific (Charles, 1986). The existence of a police personality has yet to be empirically validated. In terms of personality, law enforcement officers are a very heterogeneous group (Elam, 1983). However, there are many studies that together provide information about several personality features often found among both successful and unsuccessful police officers. The second line of research, according to Bartol (2006), was the effort to find psychological instruments that could select-in as well as screen-out suitable police candidates.

In regards to the selecting-in approach, organizations are constantly striving to attract the “right type of person.” Lefkowitz (1977) claimed that the identification of potential predictors of police officer success is essential for selection purposes. The concept of “good character” is an interesting one and there are many historical accounts of what makes a good character for purposes of law enforcement work (Ainsworth, 1995). According to Blau (1994), police managers identify the “good cop” as having the following characteristics: Bravery or courage, decisiveness, consistency and reliability, resistance to stress, cooperativeness, traditional values, and respect for authority. There is
limited research on the behavioral traits of the “best” officer. The “right personality” is often considered by police recruiters as one of those “you know it when you see it” phenomena (Ainsworth, 1995). However, such subjective judgments often preclude scientific inquiry and most importantly validation.

Recent research has demonstrated that there are particular personality factors that can be linked to police officers’ success in carrying out the essential features of the job (Cuttler & Muchinsky, 2006; Aamodt, 2004; Ones, Viswesvaran, Cullen, Drees, & Langkamp, 2003; Schneider, 2002; Sarchione, Cuttler, Mucinsky, & Nelson-Gray, 1998; Wells, 1991). Fenster and Lock (1973) pointed out that research has identified emotional stability as a critical factor in determining the probability of success in law enforcement. The link between emotional stability and police officer performance has been empirically established (Schneider, 2002; Black, 2000). Baehr et al. (1968) conducted a study on patrolman performance within the Chicago Police Department and concluded that “the ideal attributes for success are all related to stability…” (p. 231). The authors claimed that although the results are drawn specifically from the Chicago Police Department, the findings may be generalizable to other urban police departments. Both Schneider (2002) and Black (2000) found a relationship between personality test scales that represent the construct of emotional stability and officer suitability. Although research exists indicating that certain character features are related to police officer success on the job, there is not significant evidence at present to support the validity of any instrument or combination of instruments in the selection of the “best officer” amongst an applicant pool of other qualified individuals. To date, there is not a clear profile of the “best” police officer.
Once law enforcement managers began to realize the utility of psychological testing they started to rely on psychologists to identify those officers that appeared emotionally unstable rather than those who would be most successful (Reiser, 1982b). The screen-out approach to selection has been far more successful than the select-in approach in that scientists have been able to link poor performance of law enforcement officers to a number of indicators on popular personality inventories (Bartol, 1991). According to Super and Crites (1949), personality factors are a key determinant in the way people behave in any situation including how they respond to situations at work. Therefore, problems with personality functioning will likely translate to performance problems on the job. The authors conclude that, for this reason, personality assessment is such an important aspect of personnel screening and that the use of personality measures can aid in screening out applicants that evidence character features likely to interfere with the successful performance of essential job tasks. To date, most psychologists conducting preemployment evaluations for law enforcement agencies follow the practice of “selecting out negative traits to help law enforcement departments avoid problems that result from hiring officers with personality characteristics that are antithetical to good policing” (Blau, 1994, p. 111).

Whether or not a particular constellation of character features exists that reflects the typical police officer, or whether personality testing works best to select-in the best or screen-out the worst police officer applicants, personality measures can be helpful in predicting future job performance in police officers (Varela, et al., 2004; Aamodt, 2004; Inwald & Shusman, 1984; Bartel, 1982; Marsh, 1962; Fraser, 1949). Both personality and situational factors interact to bring about behavior. However, personality differences may
be more powerful when a situation is more ambiguous and the individual must rely on their own disposition as a determinant for action (Carlson and Singer, 1975), which is often the case in police work. Police officers have a great deal of discretion and must frequently make rapid judgments about what course of action to take (Carlson, 1975). The job of police officer has much latitude, meaning that there is a high degree of discretionary judgment. Jobs with latitude make personality factors particularly relevant (Thomas, personal communication, 2007). Research supports the proposition that certain people are unsuitable for certain occupations. Personality factors are considered important in the identification of those individuals who are not a good fit for a particular job (Sellbom, Fischler, & Ben-Porath (2008; Cuttler & Muchinsky, 2006; Arrigo & Claussen; 2003; Sarchione, Cuttler, Muchinsky, & Nelson-Grey, 1998; Roe, 1956). Those same character features would seem to be the best predictors of which applicants are more or less suitable for the job of law enforcement officer.

Past selection strategies have focused on eliminating unqualified individuals from the applicant pool based on biographical, physical, and character criterion (Chandler, 1990; James et al., 1984). Blau (1994) noted that police psychologists have reinforced this approach by engaging in such practices, probably because of their clinical training, which is focused on assessing psychopathology. According to James et al. (1984), the focus on screening out unsuitable applicants has narrowed researchers’ attention to characteristics associated with poor police performance at the expense of those personological variables that are associated with successful performance.
Preemployment Psychological Evaluations of Law Enforcement Candidates

Personnel selection is one of the most controversial, time consuming, and costly issues faced by law enforcement administrators (Cochrane, Tett, & Vandecreek, 2006; Colarelli and Siegel, 1964). The proper screening and selection of law enforcement officers is becoming increasingly critical (Rybicki & Nutter, 2002; Ainsworth, 1995; McCriddy, 1974). Law enforcement officers are given more power than any other professional in our society. They are legally permitted to question, search, and use lethal force if necessary. Due to the sensitive and high-risk nature of the position, it seems obvious that police applicants should not be encumbered by medical or psychological conditions that will interfere with the safe and effective performance of their duties. In most states, both medial and psychological examinations are a legal requirement. Beyond the need for police officers to be relatively free from impairing psychological disorder relative to the job, it is important that those individuals who are given such power are reliable; that is, not prone to error (Hilber & Kurke, 1995). The value of psychological testing in selecting police officers has long been recognized (Dantzker & McCoy, 2006; Super, 2006; Janik, 1994). However, the validation of these tests has been met with less enthusiasm (Baehr et al., 1968).

Major advances have been made in terms of communications systems and non-lethal weaponry. Unfortunately, the advances in the physical sciences have not been matched by those in the behavioral sciences (Baehr et al., 1968). The authors suggest that “better police officers are of greater importance than police tools for improved law enforcement.” (p. 223). According to Abbatiello (1969):

In view of the importance of the law enforcement function in our society, and the
great authority and responsibility vested in the individual police officer, it seems strange that the contribution of behavioral sciences in identifying, selecting, training, and placing the [women and] men doing the job have generally not equaled the effort expended on improving police technology and weaponry (p. 44).

Despite the emphasis of law enforcement agencies on developing tools for policing rather than screening those who will discharge the duties of police officer, the selection of law enforcement officers has progressed significantly since 19th century England.

Cochrane et al. (2003) reported that police departments serving larger cities employ extensive and complex selection systems. The personnel selection process requires systematically “collecting and evaluating information about individuals in order to extend an offer of employment” (Gatewood & Field, 1994, p. 3). Empirically validated tests for purposes of preemployment screening are immensely valuable and more valuable if validated for the specific situation in which it is being used (Kurke & Scrivner, 1995; Gowan & Gatewood, 1995; Hartman, 1987; Cronbach, 1949). According to Baehr, Furcon and Froemel (1968), “The use of psychological tests without appropriate validation research is not only unwarranted, but may be dangerously misleading both to the test user and to the test respondent” (p. 2). Horstman (1976) contended that organizations without validated employment selection technologies must either validate or cease such practices altogether, thus leaving selection to chance.

According to Gowan and Gatewood (1995), designing a selection program is usually the job of human resource specialists. However, it is recommended that these specialists work with other professional who have knowledge of law enforcement work and training in selection techniques in order to ensure that the selection system measures what it intends to measure, which is job related knowledge, skills, and abilities. Horstman (1976) as well as Bartol and Bartol (2006) reported that the validation of employment
selection systems has been a major emphasis of Industrial and Organizational Psychology since the 1970s. Additionally, law enforcement agencies are increasingly relying on psychologists with specialty training in forensic psychology. These professionals have unique training in psychological assessment as well as training in legal matters related to the application of testing.

Legal Issues

Prior to the litigation and legislation that occurred during the latter part of the 20th century, employment decisions were frequently based on factors unrelated to individuals’ ability to perform the job (Hibler & Kurke, 1995). One consequence of the litigation was the mandate to make hiring decisions based on bona fide occupational requirements (Equal Employment Opportunity Commission, 1978, 1979; Society for Industrial and Organizational Psychology, 1987). If a hiring procedure is discriminatory against a member of a protected class, it may nevertheless be defensible if it is “job related and consistent with business necessity” (Flanagan, 1995, p. 107). Therefore, screening activities employed for purposes of preemployment selection should be established as “job related and consistent with business necessity.” The formal establishment of job relatedness occurs through the job analysis (Levy, 2006). A properly conducted job analysis is the systematic investigation and description of the job and the formal establishment of the requisite knowledge, skills, abilities, and attributes for successful job task performance (Borum, Super, & Rand, 2003). A detailed description of the job analysis is beyond the scope of this paper. More information about the job analysis can be found in Levy (2006); Borum, Super, and Rand (2003); Gowan and Gatewood (1995); and Hibler and Kurke (1995).
Preemployment screening is increasingly affected by legislation and judicial decision making (Bartol, 2006). The selection process is performed under legal and environmental constraints to protect the future interests of the organization and the individual (Gatewood & Field, 1994). Personnel selection was once a cost-benefit issue. Currently, test validation is seen as a legal necessity. The shift of professionally developed and validated selection systems from organizational nicety to business necessity came with the passage of the Civil Rights Act of 1964 (Baehr et al, 1968). The Equal Employment Opportunity Commission was developed to enforce Title VII of the 1964 Civil Rights Act. The EEOC adopted the American Psychological Association’s Standards for Educational and Psychological Testing (1999) “as the model for validation of selection tests” (Horstman, 1976). The decision made by the EEOC took psychological ethics and made them the legal standard in employment selection practices. According to Horstman (1976), the EEOC’s decision made the work of test validation in employment selection the job of industrial and organizational psychologists. In addition to the legal issues related specifically to test validation, there are several regulations and standards that guide the practice of psychological evaluation of high risk occupations such as law enforcement officers.

According to Borum, Super, and Rand (2003), the Americans with Disabilities Act of 1990 (ADA, 1991) was “one of the most significant and far-reaching legal provisions affecting these assessments” (p. 135). The Americans with Disabilities Act (ADA) of 1990 directly impacts preemployment psychological evaluations because it prohibits employers from discriminating against individuals with disabilities that can perform the essential functions of a job with or without reasonable accommodations.
Specifically, the ADA prohibits any inquiries or examinations that could reveal information about an individual’s past or present medical condition until after a conditional offer of employment has been made. Before the ADA prohibited employment discrimination, many agencies required medical and psychological evaluations prior to being given a conditional offer of employment. As a result of the ADA, no disability related inquiries may be made until the applicant has received a conditional job offer. The rationale for this law is to ensure that individuals who are otherwise qualified for a given position are not discriminated against based on disability (Hibler & Kurke, 1995).

Another law developed to prevent discrimination in employment selection is the most recent version of the Civil Rights Act (CRA, 1991). The revised CRA was adopted to prevent discrimination based on gender, race, or creed. One way in which the CRA prevents employment discrimination is by restricting the use of differential cutting scores based on race. Many tests provide scores based on different comparison groups (e.g., age, sex, and ethnicity) and thus violate CRA requirements. Such practices can be avoided through mathematical correction, which combines the normative data for the different comparison groups. Another way to avoid such violations is to develop local norms (Ben-Porath, 2008).

In addition to the various federal regulations, there are a number of court cases that have significantly influenced the practice of personnel selection, including preemployment psychological evaluation of police officer applicants. The 1970 Supreme Court ruling in Griggs v. Duke Power Company had a significant impact on employment selection in the United States. In this case, the court declared that any test that discriminates on the basis of race is unlawful if the test is not directly related to the job.
Several cases thereafter expanded on the 1970 Supreme Court ruling providing specific criteria for determining both discrimination and if a test is “job related” (Horstman, 1976).

In *Soroka et al. v. Dayton Hudson Corporation* (1991), the California appeals court ruled that “invasive psychological tests violated both the constitutional right to privacy and statutory prohibitions against improper inquiries into a person’s sexual orientation and religious beliefs” (Bartol, 2006). However, the court distinguished between the use of tests such as the MMPI and the CPI for screening public safety personnel versus store security personnel. Soroka et al. established that such testing and the subsequent invasion of privacy are warranted for evaluating individuals applying for positions in the public safety arena. However, the tests must be given at the proper time and never as an entry screening tool.

In *Leonel et al. v. American Airlines, Inc.* (2005), the United States Court of Appeals for the Ninth Circuit ruled that American Airlines violated the plaintiff-appellants Federal rights under ADA by not completing all relevant non-medical components of the preemployment examination before making a “real” job offer. The ADA requires that the non-medical and medical aspects of the selection process be separated, the medical examination being the second step, and that a job offer is real only if all relevant non-medical information has been obtained and analyzed. The reason for the bifurcated selection process is so that applicants can know if they were rejected because of disability or because they lacked sufficient knowledge, skills, abilities or attributes necessary to safely discharge the duties of the position. The bifurcation process also serves to prevent the invasion of privacy until less privacy-invasive (e.g., non-
medical) inquiries have been made. American did not fulfill that legal requirement thus violating the appellants federally protected rights. American rescinded the conditional job offer and by not adhering to the ADA, it cannot be distinguished whether or not the appellants’ were terminated for medical or other reasons.

The practice of psychological evaluation, especially psychological testing, for purposes of employment selection is extremely regulated and highly restrictive. Despite the many federal, state, and local dictates on employment selection, there are several court cases that have upheld the right of law enforcement agencies to conduct psychological testing (Bonsignore v. The City of New York, 1981; McKenna v. Fargo, 1987; Conte v. Horcher, 1977). In addition to the landmark legal decision, there are several additional decision that not only support police departments’ use of psychological testing in selecting officers, but hold them accountable when such precautions are neglected in the hiring for high risk positions.

In Bonsignore v. City of New York (1982), the court ruled in favor of the plaintiff rewarding a large settlement citing negligence on the part of the police department for not requiring psychological screening of police officer applicants. The Court decided that had the agency implemented psychological testing, which was reasonably within their means, they would have found that Bonsignore was mentally ill and unfit to carry a fire arm. The central issue in the Bonsignore case was that of vicarious liability. The Court found that law enforcement agencies can be held responsible for the actions of their employees. The landmark decision resulted in the requirement of employers to show that they have taken reasonable precautions in selecting (and retaining) applicants who are relatively free from psychological problems that would interfere with safely carrying out the duties of police
officer. Bonsignore led to the NYPD adopting a comprehensive psychological evaluation process for its police officer applicants.

In Conte v. Horcher (1977) the court upheld the right for law enforcement agencies to conduct psychological evaluations for incumbent officers. This case involved a lieutenant who was ordered by the police chief to undergo psychological testing after using excessive force in the line of duty. The central issue in Conte was whether or not the police chief had the power to order the evaluation and ultimately whether or not the order was valid. The courts ruled that the police chief or superintendent does indeed have the power to order a medical and/or psychological examination in order remain informed of officers’ ability to perform the necessary job functions. However, the authority to order an evaluation is limited to those situations where ensuring the effective performance of the department is at issue (Flanagan, 1986). Although Conte involved an incumbent officer, it affirms the other courts opinions regarding the responsibility of law enforcement agencies to ensure that those they employ are suitable and stable for the high risk occupation of law enforcement officer.

In McKenna v. Fargo (1987), firefighter applicants with Jersey City challenged the constitutionality of the City’s preemployment psychological testing requirements. The City asserted that the psychological evaluation was required in order to determine whether or not an applicant was able to endure the psychological demands inherent in the job. The District Court acknowledged that some of the test questions were related to political and religious beliefs. However, the Court held that because the purpose of the test was to assess for psychopathology and not to measure orthodoxy of beliefs, the test was not an infringement of firefighters’ First or Fourteenth Amendment rights. The Court
handed the right of the City to mandate firefighter applicants to undergo psychological testing stating that the interest of the City in screening out applicants who were not able to withstand the stress of the job was sufficient to justify the intrusion on privacy. The McKenna decision was echoed in the Soroka case, reaffirming the importance of public interest (e.g., safety) over individual privacy rights and allowing public safety agencies to require psychological evaluations of candidates.

According to Flanagan (1986), there are three additional legal decisions relevant to the practice of psychological screening of police officer applicants. First, in McCabe v. Hoberman (1969), a police officer applicant who was rejected based on disordered personality, later obtained two different expert opinions stating that he was suitable for the job of law enforcement officer. The court decided that it was the agency’s prerogative which opinion to accept so long as the department acted reasonably and responsibly in coming to the decision. Second, in Peluso v. Gourdine (1982), another police officer applicant was rejected on the basis of unsuitable character after being diagnosed with features of various personality disorders. The rationale for rejecting the applicant was that he was too fragile for the stressors inherent to the job. The applicant obtained two additional evaluations stating that he was suitable. The Court affirmed the police department’s rejection stating that the agency’s decision to refuse hiring based on a disability was not illegal discrimination if the handicap interfered with the individual’s ability to safely and effectively carry out the essential job functions of police officer. The last case involved an officer named Bartucca who filed suit against the city of New Rochelle in New York. Bartucca was fired after being found unsuitable for police work based on the results of the MMPI. The New York State Supreme Court affirmed the
decision of the agency and stated that the police department did not need to rehire Bartucca. The rationale for the decision was that jurisdictional interference regarding the use of such a widely accepted instrument was inappropriate when that instrument is administered in a standard and fair manner. As stated previously, the legal decisions allowing and requiring psychological testing for purposes of preemployment screening are many. However, many agencies are out of compliance with regulations, and do not abide by recommendations or heed relevant case law.

Assessment Procedures

Despite the use of sophisticated selection systems, many of the law enforcement agencies serving larger cities do not adhere to public policy guidelines and professional standard for psychological assessment (Cochrane et al., 2003). According to Borum, Super, and Rand (2003), current guidelines and practice standards for preemployment psychological screening require the use of objective psychological testing and a job related interview. A face-to-face interview that is structured and behaviorally oriented should always be conducted as one part of the screening process (Specialty Guidelines for Forensic Psychologists, 1991; Hartman, 1987). Additionally, research should exist on the validity of those tests used in preemployment selection (Principles for the Validation and Use of Personnel Selection Procedures, 2003; Hargrave & Berner, 1984). Lastly, federal law dictates what type of tests can be administered at different phases of the selection process: The legal prohibition on the use of tests that are medical in nature prior to the COE is one example (Leonel et al. v. American Airlines, Inc., 2005).

Strawbridge and Strawbridge (1990) conducted a survey on selection procedures for large law enforcement agencies and found no standard practices for psychological
evaluations. The authors also found that practices varied greatly across agencies in terms of what tests are administered and how the interview is conducted. Leake (1988) reported that less than two percent of law enforcement agencies in California engage in research investigating personal variables related to successful and unsuccessful job performance of officers. These agencies also reported no intention to conduct such research in the future.

Currently, test validation is seen as a legal necessity. The federal government, through the publication of Guidelines on Employment Testing Procedures (EEOC, 1966), essentially set a standard for validating test prior to their use in selection decisions. Such a lack of adherence to professional standards, as well as a disinterest in federally mandated test validation, is cause for concern. Non-compliance is likely a result of a lack of knowledge regarding the constantly evolving and complex legal contours of personnel selection rather than intentional disregard. If ignorance is the case, communication and education at both the administrative level as well as the individual level for those who are carrying out the various selection procedures, should resolve the issue.

Currently, many states require psychological testing in the selection of law enforcement officers (Janik, 1994). Some states go so far as to dictate which tests are to be used. According to Elam (1983), the state of Oklahoma requires that all police candidates be evaluated by a measure similar to the MMPI; the CPI was considered to be an equivalent measure. Blau (1994) reported that the Los Angeles Police Department requires psychological and psychiatric screening for all police officer candidates. Blau noted that the LAPD selection system consisted of a civil service exam, a face-to-face interview, a background investigation, and a physical examination. The
psychological/psychiatric evaluation consisted of the MMPI, other personality tests, and a brief interview.

According to Barehens (1985), many states mandate police officer screening by statute. According to Baehr et al. (1968), Chicago’s police officers must pass an extensive screening process before they are hired. Based on the extensive pre-screening process undertaken by the Chicago Police Department, tests assessing intellectual functioning, such as the WAIS, were deemed unnecessary because the Civil Service Examination was assumed to screen out individuals with below average intellectual functioning (p. 57). The Chicago Police Department also uses tests of attitude and personality (Baehr et al., 1968) to screen applicants. At present, the New York Police Department and the LAPD require both psychological and physical examination of officer candidates. Most agencies have the psychological and physical examination as requirements for employment screening (Janik, 1994).

According to Hartman (1987), “Psychological screening of law enforcement candidates has become the norm rather than the exception” (p. 5). At present, preemployment psychological evaluation is standard practice among US police agencies (Cochrane et al., 2003). A survey of federal, state, and local law enforcement agencies in the southeastern United States revealed that over half of the agencies employed psychologists who used both a clinical interview and an objective personality measure (Super, 2006). A nationwide survey of psychological screening trends conducted by Behrens (1985) revealed that 50 percent of responding agencies implement psychological screening in the selection process. The majority of those agencies began using psychological screening after 1981 (Behrens, 1985). According to Ainsworth (1995),
approximately 70 percent of American agencies utilize psychological testing. Another study by Rybicki and Nutter (2002) showed that most agencies, small, medium, and large, employ preemployment psychological evaluations of police applicants whether the screening method is done through the department or through outside contract psychologists.

Interestingly, the British government recommends against the use of psychological tests in law enforcement selection stating that there is not a psychological test currently suitable for such an endeavor. Cochrane, Tett, and Vandecreek (2003) conducted a national survey of 155 municipal police departments. The investigators found that the majority of large city police departments perform a background investigation, medical examination, interview, drug test, physical fitness exam, and polygraph test. The authors also found that more than 90 percent of police departments require applicants to undergo a psychological evaluation. Cochrane et al. (2003) report a shift in practices among agencies over the past decade. The results evidence a drastic increase in law enforcement agency’s use of psychological evaluations compared to past practices.

Personality tests are the most frequently used psychological measure for the purpose of preemployment screening (Hancock & McClung, 1987). Dietrich and Berger (1978) noted that the MMPI is frequently used in police selection. Inwald (1987) also reported that the MMPI is the most commonly used test in law enforcement screening. More recently, Lee (2006) found that the MMPI was the instrument of choice for employment screening of police officer applicants. According to O’Connor (1962), the most frequently used test in the 60s was the MMPI. At that time, the CPI was not used by
any of the reporting cities (Elam, 1983). Murphy (1972) found that of those agencies who utilize psychological testing in their screening and selection procedures, 50 percent indicated that they utilized the MMPI. According to Hartman (1987), the majority of law enforcement agencies use the MMPI and a clinical interview in combination with one or more of the following tests: The California Psychological Inventory, the Sixteen Personality Factors Test, the Edwards Personality Preference Schedule, and the Inwald Personality Inventory. This pattern of personality test use is true currently for the majority of law enforcement agencies (Super, 2006; Dantzker & McCoy, 2006; Cochrane et al., 2003; Scrivner, 1994).

Surveys indicate that the MMPI and the CPI are the most frequently utilized psychological assessment instruments for employment screening in law enforcement (Super, 2006; Cochrane et al., 2003; Johnson, 1983; Murphy, 1972; Poland, 1978). However, unlike the CPI, the MMPI was developed to assess for psychopathology and was normed on psychiatric patients; it is considered medical in nature and cannot be used at the pre-offer phase (i.e., prior to a conditional offer of employment) of the employment selection process. The CPI was developed for the purpose of assessing more normative personality traits and normed on a non-pathological population. Varela et al. (2004) suggested that using measures that assess normative traits may prove most effective in the psychological screening of law enforcement candidates because most candidates are exposed to several other screening phases before they reach the psychological examination and so the more pathological applicants may already be screened out. Therefore, because the CPI is not only legally permitted prior to a conditional offer of employment, but it appears to be particularly suited for assessing job candidates, it is
thought to be more valuable than tests that measure psychothathology (e.g., MMPI).

Many other issues involving reliability, validity, and inconsistent findings remain
(Costello and Schoenfeld, 1981; Inwald and Shusman, 1984a; Inwald and Shusman,
1984b; Merian et al., 1980). Nevertheless, police psychologists continue to employ these
instruments despite the unresolved issues.

According to the IACP Police Psychological Service Section’s Pre-employment
Psychological Service Guidelines (2004), a face-to-face interview is recommended as one
part of the employment screening process. Research has demonstrated that most law
enforcement agencies employ a clinical interview (Super, 2006). There is very little
empirical evidence supporting the use of the clinical interview as a predictor of job
performance (Smelson, 1975; Morris, 1979; Shapiro, 1981). Predictions based on human
judgment, such as preemployment interviews, have been found to be less accurate than
judgments based on actuarial methods (Dawes, Faust, & Meehl, 1989; Inwald, 1988;
Inwald & Knatz, 1988). However, according to Hibler and Kurke (1995), “Various types
of biographical information have different predictive values” (p. 67).

Sharf (1994) found the following categories of personal history items predictive
of successful job performance: Demographic classifiers, habits and attitudes, health,
human relations, money management, developmental information, socioeconomic data,
financial status, social activities, associates, personal attributes, home life, recreation,
interests, educational history, self-impressions, values, work history and skills. Although
most of this information could be obtained through a clinical interview, Hartman (1987)
reported that the clinical interview alone “is not a defensible technique in making job
performance predictions” (p. 5). Nevertheless, the clinical interview can provide valuable
information in the way of confirming and clarifying test data (Hartman, 1987). Meloy (2008) claimed that the clinical interview can provide important information beyond that which can be obtained through actuarial methods. Additionally, the clinical interview is recommended as one element of the comprehensive battery in psycho-legal evaluations (Specialty Guidelines for Forensic Psychologists, 1991). Lastly, according to Blau (1994), the Accrediting Standards of the Commission on Accreditation for Law Enforcement Agencies (CLEA) “includes a requirement that a clinical psychologist conduct an interview as part of the psychological testing procedure” (p. 86).

Even though no selection system is perfect, psychological testing appear to be the best insurance in sight. McDonough and Monahan (1975) contend that no screening instrument alone can predict who will be successful and who will not. Barnabas (1948) proposed that if an instrument performs better than chance, it has some utility. According to Inwald, Hurwitz, and Kaufman (1991), objective measures of personality reduce the uncertainty associated with hiring unknown job candidates. Bartol (2006) predicted that preemployment psychological screening will continue to play a crucial role in controlling counterproductive work behavior in police officers and that this task will be carried out by psychologists with special training in law enforcement issues.

Finding the Best Approach for Hitting a Moving Target

The job of police officer and the criteria for successful performance continues to evolve requiring rigorous and ongoing selection research (Wilson & Grant, 1998; Ainsworth, 1995; Blau, 1994; Azen, Snibbe, & Montgomery, 1973). Police officers are required to perform a wide variety of functions (Bartol & Bartol, 2004; Koper, 2004; Crosby, 1979) and for that reason, the role of police officer is difficult to define. The
characteristics required to survive in law enforcement are likely to vary between regions, communities, and over time (Ainsworth, 2002; Pugh, 1985; Daley, 1978). Both Varela (2000) and Levy (1967) noted that the degree of variation between agencies likely contributes to the difficulty in finding a model for successful law enforcement officers. Elam (1983) found support for this position in that the author discovered little overlap among personality tests when used as predictors of officer success in two different types of agencies, municipal and state. One particular test may be valid for one particular group in one particular situation, but most tests are not valid for most groups in most situations (Eisenberg & Reinke, 1973). According to Hartman (1987) no one test battery has been found superior to another in the psychological screening of law enforcement officers. The best method is to use several different measures in order to get multiple data points (Hogan, Hogan, & Roberts, 1996) to use for verification of candidates’ self-report.

Baehr et al. (1968) claimed that the validation of a given tests battery “could be improved by the construction of specialized tests for police officer selection” (p. 57). Tests should not only be validated on law enforcement officers as a group, but on officer samples from the specific agencies for which the test will be employed (Hartman, 1987). Hartman (1987) also notes that it is in the best interest of an agency to develop its own norms, especially in the case where the agency represents personnel dissimilar to the test’s normative sample. In so doing, researchers and practitioners will have valid information regarding successful and unsuccessful officer characteristics relative to the exact job tasks required within the particular agency for which applicants are applying. Such precision should yield more accurate and ultimately more useful information for purposes of hiring decisions. Blau (1994) suggests that until comprehensive batteries are
available that are appropriate for local conditions, psychologists would do best by adhering to national practice standards such as the American Psychological Association, the American Educational Research Association, and the National Council in Measurement of Education.

There are many questions still unanswered in the realm of police officer selection. Questions such as what type of people are interested in law enforcement? Why are certain people interested in law enforcement? What factors determine the success of law enforcement officers? Are there instruments that can predict, with any degree of certainty, how an individual will behave in certain situations typical of law enforcement? Are there ways to identify those individuals who were once qualified, but have become unable to carry out the essential job functions of police officer or worse have become a potential threat to themselves and/or others? Some of these questions have been addressed to no avail and others have simply raised more questions. There is no doubt that police selection is a complex task. However, there are some areas of inquiry that have demonstrated potential utility, such as the identification of police officer applicants who are considered unsuitable or, rather, problematic for purposes of employment as law enforcement officers; this line of inquiry, as noted above, is referred to as screening-out.

The current study is concerned with the identification of those individuals who are found unsuitable very early in the selection process. The study was designed to examine the validity of a personality measure (CPI) in predicting which applicants will be eliminated from the applicant pool during one of the initial hurdles: the preliminary background investigation. That is, those individuals judged unsuitable for law enforcement work by veteran police investigators (e.g., opinion experts [Levy, 2006;
Baehr et al., 1968]. The value of this study can be found in the fact that identifying and screening-out unsuitable applicants early in the selection process reduces costs. In the next section, a review of the California Psychological Inventory is provided in order to orient the reader to the instrument before an examination of the existing research on the CPI in police selection.

An Overview of the California Psychological Inventory

Description and Purpose

The California Psychological Inventory (CPI) (Gough, 1991) is a well known and respected test of personality (Blau, 1994). The CPI is a measure of normal personality (Gough, 1965) and is “notable for its lack of symptom-oriented material” (Magargee, 1972, p. 5). The CPI was normed on and meant to be used with non-psychiatrically disturbed individuals (Gough, 1975); it is not considered medical in nature. The instrument differs from other personality inventories such as the 16-PF and the MMPI in its simplicity (Magargee, 1972). The CPI measures everyday features of interpersonal behavior that most everyone is familiar with (Gough, 2000).

The purpose of the CPI can be found in the original intention of its developer, Harrison Gough. Gough had the intention of creating a personality measure for assessing nonclinical populations using concepts to describe interpersonal style, behavior patterns, and personality characteristics currently existing in everyday language (Craig, 1999). These concepts were thought to be universal and Gough (2000) referred to them as “folk concepts.” Gough contended that tests should be useful and useable (Magargee, 1972). The CPI items were written with the idea in mind that if simple, common, everyday
language was used, than people would be able to more easily describe themselves in familiar terms (Gough, 1987). The goal of the CPI is to “assess individuals by means of variables and concepts that ordinary people use in their daily lives to understand, classify, and predict their own behavior and that of others” (Gough & Bradley, 1996, p. 1).

There are several advantages to using folk concepts for scaling. The first advantage is that folk concepts are cross-culturally relevant. The second advantage of folk concept scaling is the ease of interpretation because of the straightforward meaning of the scales. The third advantage is the power that such variables have in describing consistent patterns of behavior and the subsequent value of those concepts in predicting future behavior (Gough, 1968).

While most of the previously discussed psychological instruments have been developed for use in particular settings or for very specific reasons, the CPI was created for large-scale application. The CPI was created with the goal of developing “descriptive concepts” that have wide-ranging relevance to individuals themselves and society as a whole (Gough, 1975). Additionally, the test developer aimed to create scales that were brief, accurate, and dependable in the identification and measurement of the concepts. The instruments scales are concerned with variables relevant to everyday life. These “folk concepts” are thought to be relevant in the understanding and prediction of interpersonal behavior across settings. Although, the CPI has been found particularly useful in certain areas (e.g. detecting and predicting antisocial behavior), it also has been shown to have utility in the educational and organizational arena (Gough, 1975).

The most recent version of the CPI is the 434 Form, which retains the original intent of its predecessors. The primary purpose of the 434 Form of the CPI is “to furnish
information to the interpreter from which a veridical (true-to-life) and useful picture may be drawn of the person taking the test. The portrait should be recognizable as accurate by friends and acquaintances, and should also provide a good starting point for predicting future behavior and for understanding prior actions” (Gough & Bradley, 1996, p. 1). A brief history of the CPI is necessary in order to fully understand the theoretical and scientific underpinnings of the instrument, its evolution, and the various applications of the instrument in the present day, including its role in the present study.

History and Development

Gough (1965) claimed that there are three ways to go about selecting traits to assess personality. The first strategy consists of relying on past psychological theory for the selection of traits to measure. According to Megargee (1972), “Gough never adopted a formal theoretical position” (p. 11). The second method, similar to that used by Raymond Catell in developing the Sixteen Personality Factor Questionnaire, involves creating a test from scratch. However, this approach may result in tests that lack utility because they are not relevant. The third and last approach, the method that Gough preferred, involved using concepts that already exist within the setting for which the test is to be used. One example of the third approach is that of the Strong Vocational Inventory, which was created for occupational guidance purposes and developed using concepts found in the workplace.

Gough developed the CPI, a measure of normal personality, using terminology that people in everyday life use to describe and characterize one another’s behavior and interpersonal style. Additionally, Gough required that the terminology stood the test of time and was not limited to any one place. The result is what Gough calls folk concepts:
variables such as responsibility, tolerance, and sociability that can be found in the vocabulary of most all cultures that use symbols to communicate (Magargee, 1972).

The CPI was meant to serve a pragmatic, or rather “instrumental” function. Unlike the definitional purpose of intelligence tests where the test itself defines the construct, the value of the CPI is determined based on its accuracy in classifying people as they would be classified by others and in its ability to accurately predict behavior in particular contexts (Gough, 1987). Because the CPI was born from this perspective, the test developers focused less on “psychometric eloquence” (Groth-Marnat, 2003, p. 355) and more on practical utility (Gough, 2000). Gough (1987) was more concerned that the test assess the complex of qualities as they manifest “in the folk” rather than developing a test with superficial accuracy that adhered to statistical tradition such as orthogonality among scales, which defies the intercorrelational nature of interpersonal appraisal from a social psychology perspective.

Gough (1987) said that the CPI has only two basic aims, which are 1) “to predict what people will say and do in specified contexts,” and 2) “to identify individuals who will be evaluated and described in particular and interpersonally significant ways” (p. 4). Gough (1987) makes the intention of his test very clear when specifying that the CPI was not intended to define or assess psychological traits. But rather, the CPI was developed to classify people as they would be described by others (e.g., folk concepts).

Gough attempted to create a measure of certain character dimensions using the MMPI and actually developed the MMPI-derived scales of Social Status, Prejudice, Dominance, and Responsibility. However, the MMPI consisted of mostly pathology-
based items, which precluded the goal of developing an instrument for nonclinical populations (Gough, 1987).

Gough (1987) reported that the development of the CPI was well underway by the late 1940s with two scales developed (e.g., Capacity for Status and Tolerance) and three under construction (Dominance, Responsibility, and Socialization). Gough published the first scales of the CPI in 1948. After further research and development by Gough and colleagues, the CPI was then compiled into a single booklet. The first copyrighted edition, a 15-scale inventory, was published in 1951 (Magargee, 1972).

In 1951 the CPI was released in its initial 548-item form, which scored 15 of the 18 scales later published in the 1956 480-item version. After publishing the original 15 scales, Gough added three scales to the inventory (Sp, Sa, and Sc), publishing the 18 scale measure. In 1956 the full 18-scale inventory was released by Consulting Psychologists Press (Magargee, 1972). The CPI was first reviewed by Laurance Shaffer in 1957.

The inventory was later revised in 1987 (Groth-Marnat, 2003). By the time the 1987 manual was released, 18 items had been dropped: twelve of which were simply redundant and 29 items were reworded to 1) reflect a more current phraseology, 2) be less sex biased, and 3) be more easily read and understood. The 1987 version also included two new scales (Independence and Empathy), which reflects the current 20 folk concept scale inventory (Gough, 1987). At that time, the 20 scales were “intended to be sufficient to permit explication and prediction of a broad range of interpersonal behavior” (Gough, 1987, pp.1-2). However, Gough (1965) did not consider his scales to be inclusive of all
folk concepts or even most. According to Gough (1987), there is also empirical support suggesting that the CPI is valid in the prediction of both academic and job performance.

The CPI was most recently revised in 1996 (Craig, 1999). The measure was restandardized using 3,000 men and 3,000 women (Gough & Bradley, 1996). Form 434 contains 28 fewer items than the previous form, which were thought to be in conflict with fair employment practices according to the 1991 Americans with Disabilities Act or in possible violation of certain privacy rights. Additionally, some of the items were omitted because a significant number of individuals found them to be objectionable (Gough & Bradley, 1996). Lastly, there is also a 250-item short form that is currently under development (Groth-Marnat, 2003).

**Scales and Vectors of the California Psychological Inventory**

The CPI items tap information regarding an individual’s characteristic behavior as well as feelings and attitudes concerning social matters (Groth-Marnat, 2003). The CPI scales are designed to forecast what a person will say or do under defined conditions, and to identify individuals who will be described in characteristic ways by others who know them well or who observe their behavior in particular contexts (Gough & Bradley, 1996). The results of the test are plotted on 20 scales and 3 vectors (factors) that reflect descriptions of social relationships in common parlance. The scales are intended to 1) predict what people will say and do in specified contexts, and 2) identify individuals who will be evaluated and described in differentiating and interpersonally significant ways (Gough & Bradley, 1996). The purposes of the CPI are different from other personality instruments in that no claim is made regarding the definition and assessment of
psychological traits. The scales were developed free of any linkage to trait notions (Gough & Bradley, 1996).

Many authors refer to the CPI as being a “prototypic example of empirical methodology” (Gough, 1987, p.10) because the development of CPI scales is based on two different approaches: The empirical method and the internal consistency method. The empirical method refers to the analysis of items against nontest criterion and the subsequent selection and keying of items in a way that maximizes the relationship between responses and predictor (Gough, 1987). The internal consistency method involves selecting items thought to be related to the purpose of the test, then analyzing the intercorrelations among the items to omit those items that are least consistent with the larger psychometric theme (Gough & Bradley, 1996). Gough (1987) posits that the value of a scale should be found in its functional utility rather than how it was developed. In the 462-item version of the CPI, 194 items came from the MMPI. The Sociability scale was developed based on an item-analysis of MMPI protocols. Other scales were developed using a combination of items from the MMPI and newly constructed items (Gough, 2000). The Intellectual Efficiency scale was developed using a mixture of MMPI items and new items. The Good Impression scale consisted of all original items (Gough, 1987).

The CPI consists of 20 primary scales (McAllister, 1996), each intended to measure an important feature of either individual or interpersonal psychology. Together the scales provide a comprehensive picture of the individual from a “social interaction or ‘folk concept’ point of view” (Gough, 1975, p. 5). Of the 20 scales, 13 were developed empirically: Dominance (Do), Capacity for Status (Cs), Sociability (Sy), Independence (In), Empathy (Em), Responsibility (Re), Socialization (So), Tolerance (To),
Achievement via Conformance (Ac), Achievement via Independence (Ai), Intellectual Efficiency (Ie), Psychological-mindedness (Py), and Femininity/Masculinity (F/M). Four of the scales were developed using the internal consistency method: Social Presence (Sp), Self-acceptance (Sa), Self-control (Sc), and Flexibility (Fx). The three remaining scales were developed using a mixture of the two strategies: Good Impression (Gi), Communality (Cm), and Well-being (Wb) (Gough, 1987). The scales are divided into four separate domains and items are grouped based on their related implications.

According to Magargee (1972), the CPI scales are grouped into clusters or classes for ease of interpretation. Class I Scales: Measures of poise, ascendancy, self-assurance, and interpersonal adequacy. Class II Scales: Measures of responsibility, socialization, maturity, and interpersonal structuring of values. Class III Scales: Measures of intellectual efficiency and achievement potential. Class IV Scales: Measures of intellectual and interest modes. The scales are grouped for convenience into four broad categories, bringing together those having related implications. The underlying logic here is interpretational, not factorial, i.e., these four categories do not necessarily constitute psychometric entities (http://cps.nova.edu/~cpphelp/CPI.html).

Of the 20 scales, three are related to validity (faking-bad, faking-good, and popular or frequent responses) and are considered a measure of an individual’s test-taking attitude (Groth-Marnat, 2003). According to Gough (1975) the CPI also has scales that have been developed to detect deliberate dissimulation or faking. Gi (Good Impression), Wb (Well-being), and Cm (Communality) are scales that contain more subtle items and assist in detecting the deliberate exaggeration and consequent distortion of the test results. High scores on Gi are an indicator of attempts to place oneself in a favorable
light. Very low scores on the scale of Wb indicate exaggerated personal distress or rather, faking bad. Lastly, Cm is an indicator of the individual’s approach to test taking. This scale represents a common denominator across samples. Very low scores reflect atypical response patterns and suggest random responding.

Extensive research has revealed that the CPI scales can be reduced to four or five factors (Gough, 1987). Of the factorial findings, two principle themes or dimensions were best established. The first theme reflected an “outgoing, self-confident, and interpersonally interactive” disposition (e.g., introversion versus extraversion) and the second theme involved “internalization of social imperatives and the control of impulses” (e.g., norm-favoring versus norm-rejecting). Two new scales were developed from the research on the two new factors. The scales were named “Person Orientation” and “Value Orientation.” From a similar line of research, another separate factor emerged from within the two new scales. This third factor was considered a measure of self-realization or personal goal attainment (Gough, 1987) and contains three scales: Intellectual Efficiency, Tolerance, and Wellbeing.

The three factorial structures are referred to as “Vectors.” People scoring high on Vector 1 tend to be seen as “reticent, shy reserved, moderate, modest, and reluctant to initiate or take decisive social action,” those scoring high on Vector 2 tend to be viewed as “well-organized, contentious, conventional, dependable, and controlled”, and lastly, individuals scoring high on Vector 3 tend to be described as “free of neurotic trends and conflict, moderate, mature, insightful, optimistic, and as having a wide range of interests” (Gough, 1987, pp. 14-20).
The three Vectors were developed into a cuboid model where Vector 1 and Vector 2 combined to produce four possible permutations or type categories: Alpha, Beta, Gamma, and Delta. Alphas are “enterprising, dependable, and outgoing”, Betas are “reserved, responsible, and moderate”, Gammas are “adventurous, restless, and pleasure seeking”, and Deltas are “withdrawn, private, and to some extent disaffected.” Each type category also manifests a level of self-actualization ranging from level 7, the highest level of psychological integration for that type, to level 1, little to no self-realization of the type (Gough, 1987).

Gough has come to the conclusion over the years that if the relationship of individual scales is too low in a given setting, one must move to using patterns and combinations of scales (Gough and Kirk, 1970). As a result, Gough has developed assessment algorithms from the CPI scales using multiple regression techniques (Magargee, 1972). The eight risk ratings generated from the Police and Public Safety Selection Report are one example of such algorithms.

The ability for the CPI to predict successful job performance and the parameters that influence those predictions have been well-documented in the literature. According to Megargee (1972) “There is a surfeit of multiple regression formulas” (p. 251). Megargee suggests that future research focus on testing the already existing prediction formulas before developing new ones.

**Peer Review and Empirical Support**

The items of the CPI were selected as a result of empirical and theoretical methods. Although many of the questions were initially developed using a rational approach, all final items were the result of empirical criterion keying. A unique feature of
the CPI is what Gough referred to as the “open system.” In an open system, elements can be added or dropped from the system as evidence supports such alterations. As an open system, the CPI has evolved substantially from the time of its conception.

According to Groth-Marnat (2003), the majority of reviews of the CPI have been favorable. One reviewer described the measure as “an excellent normal personality assessment devise, more reliable than the manual advertises, with good normative data and outstanding interpretive information” (Bolton, 1992, p. 139). Magargee (1972) noted that Klieinmuntz (1967) claimed that the CPI was “…well on its way to becoming one of the best, if not the best, personality-measuring instruments of its kind” (p. 239). Anastasi (1968) claimed that the CPI was “one of the best personality inventories currently available” (p. 448). Anastasi (1968) also praised the developers of the inventory for their empirical rigor. More recently, Atkinson (2007) reviewed the CPI calling it “successful in its groundbreaking attempt to describe a broad array of fairly robust personality characteristics” and claiming that “the CPI can assist in the description of individual’s personal and interpersonal characteristics associated with stable dimensions of personality” (p. 4-5).

The CPI was first used in a large-scale research testing project in 1951 and has been administered to at least five million subjects (Gough, 1975). Although most of the cases have not been available to the publisher, a sufficient number have been collected to provide typical profiles for a number of different groups. According to Megargee (1972), research on the CPI has progressed rapidly from the time of its initial publication. Gough’s first manual presented 44 studies concerning the CPI. Approximately 25 years later, over 600 studies exist that used or investigated the CPI. According to Groth-Marnat
(2003), the CPI has been used in more than 2,000 research studies. Since its inception nearly five decades ago, an extensive body of research has formed providing validation for the use of the CPI in a variety of settings and providing a wealth of knowledge regarding personality as it relates to various outcomes (Atkinson, 2007). Additionally, according to Hattruo (2007), the CPI has considerable practical value in predicting a multitude of behavioral outcomes.

The test was originally normed on a sample of 1000 men and 1000 women representing the general population based on age, education, SES, and other relevant parameters. The original normative sample included 50 male police officers and 50 male prison guards (Gough, 1987). The normative sample for the 1957 inventory consisted of 6,000 males and 7,000 females varying widely in age, SES, and geographic area (Groth-Marnat, 2003). The 1996 revision was standardized on a sample of 3,000 equally distributed between men and women and matched to the US census for relevant variables (Gough & Bradley, 1996).

In the CPI manual, Gough (1975) gives a general account of the research used to validate the CPI. Test-retest studies have been used to demonstrate that the test is a highly consistent measure. Several studies measuring the validity of the CPI indicate that the individual scales are of acceptable validity. The majority of studies were concurrent validity studies resulting in a range of validity coefficients from a low of .21 on Sc (self-control) to a high of .60 on Gi (good impression). Most of the studies yielded correlations above .40 and many were closer to .50 (Gough, 1975).

The CPI has been subject to over 50 years of empirical scrutiny. As a result, the author has continued to refine and improve the instrument based on the research.
Improvements have consisted of numerous predictive validity studies, the development of alternative scales, and expanded normative data (Groth-Marnat, 2003). The extensive empirical work and subsequent refinements have been included in the most recent 1996 version of the inventory (Gough & Bradley, 1996). Gough, 2000 noted that because of the extensive research and development involved in the most recent version, the CPI has become a highly regarded and frequently used test, particularly in the areas of career development, personnel selection, interpersonal maladjustment, and predicting antisocial behavior (McAllister, 1996).

The developers of the CPI were less concerned with psychometric elegance and more interested in prediction (Groth-Marnat, 2003). As a result, much of the research on the CPI focuses on the success of the individual scales and regression equations in predicting what people will do or say in certain situations. The research has shown that the test performs favorably as a predictive instrument and is of practical value in the field of personnel selection (Megargee, 1972). In keeping with Gough’s (1968) philosophy that tests should be developed for practical purposes, his research on the CPI focused on looking at important everyday behaviors such as school performance, work performance, and creativity. Gough considered the validity to be justified if the test was successful in contributing to accurate predictions of such behavior patterns. This approach is referred to as practical validity (e.g., concurrent and predictive validity) and it is different from what has been termed trait or construct validity.

As a result of its practical usefulness, the CPI has become one of the most frequently utilized measures for psychological assessment (Camara et al., 2000). There is empirical support for the use of single scales and combinations of only a few scales for
predicting outcomes such as police officer performance (Hogan, 1971; Hogan & Kurtines, 1975). Other criteria for which the CPI was found useful in forecasting are achievement in high school, college, and certain health professions (Gough, 1987). In a personal communication between Gough and Magargee (1971), Gough claimed that the eighteen scales could predict “just about everything that happens in interpersonal life” (Megargee, 1972, p. 13). However, Gough (1965) did not claim perfect prediction and suggested that in most cases prediction is modest. Gough (1965) also claimed that prediction is enhanced by considering additional sources of information and contextual factors. Research has demonstrated the predictive validity of the CPI in areas such as academic achievement, work performance, and delinquency. However, there was limited study of the CPI in law enforcement over 30 years ago. Gough (1975) stated that any issue involving social or interpersonal behavior is a proper research concern for the CPI.

Administration, Scoring, and Interpretation

The California Psychological Inventory (CPI) is self-administered and can be delivered via paper-and-pencil or a computer. The inventory can be read by the respondent or the examiner. Items may be left blank, but completed forms are best for scoring/interpretive purposes and research (Gough, 1987).

The CPI is a forced choice test composed of 434 true-false statements. The CPI requires about a seventh grade reading level which translates to approximately 13 or 14 years of age (Gough, 1987; Gough & Bradley, 1996). No time limit is imposed although most individuals complete the test within an hour (Magargee, 1972). Typical test time ranges from 45 to 60 minutes (Gough, 1975). Testing may be divided into two or more sessions for special circumstances such as slow reading rate or time constraints (Gough,
The test can be administered to individuals or to groups. Although standard testing conditions are advised, “Standardized testing conditions are not essential” (Magargee, 1972, p. 5). The test publisher suggests that rigorous conditions need not be established (Gough, 1975). In most psychological testing endeavors, the CPI has been shown equally reliable and valid under “nearly every conceivable condition” (Gough, 1987, p. 11).

According to Gough (1975), the CPI can be scored by hand or through the use of computer scoring and interpretive software. The computer scoring service also includes an option for a computer generated interpretive report (Gough, 1987). A more recent source (Gough & Bradley, 1996) indicated that scoring the Form 434 must be done through Consulting Psychologists Press. Similar to the MMPI, raw scores are converted to standard T scores (Magargee, 1972) with a mean of 50 and a standard deviation of 10.

The scales on the profile are presented in order from the more observable social-interactional qualities (e.g., Dominance and Sociability) to the more internal qualities such as values, controls, and modes of operating respectively (e.g., Responsibility, Tolerance and Flexibility). Scales are scored so that “higher values are associated with conventionally favored standing on the variable, and lower scores with a less favorable status” (Gough, 1987, p. 5), except in the case of the femininity/Masculinity scale. Therefore, for the most part, higher scores indicate strengths whereas lower scores reflect weaknesses.

The general rule for interpretation is to start by reviewing the profile for reliability by looking at the three validity scales of Good Impression, Communality, and Well-being to ensure that the standard scores fall between 35 and 65 (Groth-Marnat, 2003). Next, the profile is compared to “groups or classifications relevant to the purpose of the analysis”
The profile classifies people in four structural themes (Gough & Bradley, 1996). Then, individual scales are reviewed for their individuating and differentiating function (McAllister, 1996).

Although the simplicity of the CPI is one of the instrument’s advantages, it can also be deceiving. Those new to using the CPI may become comfortable in its use without developing a complete understanding of the very principles from which it was developed (Magargee, 1972). This mistake can result in erroneous use of the measure. Gough (1975) claimed that the utility of any psychological test is a function of multiple factors such as the test’s comprehensiveness and adequacy, the interpreter’s background and skill level, and the user’s knowledge of and experience with the test. Gough (1968) and his critics agree that the CPI must be interpreted by a qualified professional with special training in psychology (Magargee, 1972).

Lastly, Gough (1987) said “in contemporary psychological assessment it is a distinct advantage to have a range of tools available, each deriving from its own logic and theoretical roots” (p. 1). The professional user of such tools can then pick and choose those that are more helpful and accurate in the particular setting in which the assessment is to be carried out. The idea that multiple measures be used in assessment and that tests be selected for the specific purpose and context in which they are being used permeates the literature on psychological testing and is consistent with all of the standards and guidelines in personnel selection to date. Therefore, the CPI is most appropriately utilized when addressing those test questions for which the test was developed (e.g., evaluating the psychological and behavioral tendencies of nonpathological individuals) in those settings where it has been validated on the population (e.g., educational and
occupational). Although no instrument has been found sufficient in itself to make
determinations about an applicant’s suitability for a particular job, or whether an
incumbent is appropriate for advancement or special assignment, the CPI has
demonstrated added validity to prediction when used with other devices such as the
clinical interview and a personal history questionnaire (Varela et al., 2004).

Special Applications

Originally developed as a measure of normal adult personality, the CPI was also
intended to assess the vocational and career goals of individuals in career counseling
(Blau, 1994). In addition to the more traditional uses of personality measures, the CPI has
been utilized extensively for industrial and organizational purposes (Groth-Marnat,
2003). The CPI has also been used as an appraisal instrument in the determination of
individual fitness for specific jobs and job-related activities (Blau, 1994). According to
Hartman (1987), the CPI has been used for many different purposes, one of which is the
evaluation of law enforcement officers. The CPI has been found useful in the prediction
of on-the-job behavior. It has been validated on law enforcement officers (Hortsman,

Special purpose scales have also been developed for those instances where the 20
folk scales were found insufficient for classification and prediction purposes. The special
purpose scales apply to industrial and organizational contexts (e.g., managerial potential
and work orientation) as well as educational contexts (e.g., success in graduate school).
Gough (1987, p. 2). One of the special purpose scales, the Law Enforcement Orientation
(Leo) scale, was designed to identify individuals whose beliefs and interest are well
suited for work in law enforcement. According to Groth-Marnat (2003), individuals
scoring high on Leo are optimistic, stable, ambitious, conscientious, and possess leadership ability. Additionally, individuals scoring high on Leo are likely to be honest, capable of creating a good impression, and have good interpersonal skills. These individuals are often conservative, conventional, moralistic, and hard-working (McAllister, 1996).

There are several CPI reports available at present. One of the CPI reports relevant to the current study is a special report developed for the selection of police and public safety personnel. The Police and Public Safety Selection Report “is a special report developed by Michael Roberts using normative data and interpretive information specific to the identification of suitable police and public safety applicants.” (Gough & Bradley, 1996, p. 16). According to the CPI Police and Public Safety Report Technical Manual (Roberts & Johnson, 2001), the special CPI based screening report addresses many of the practical, legal, and ethical issues faced by professional psychologists conducting preemployment psychological evaluations of applicants for positions such as police officer, fire fighter and EMT, corrections officer, communications and emergency dispatcher, and juvenile probation officer. The current study used data (e.g., Risk Ratings) generated from the normative sample used in the Police and Public Safety Selection Report. The remainder of this section will be devoted to a review of the CPI Police and Public Safety Selection Report.

The specialized CPI Police and Public Safety Report was developed from the 434-item version of the CPI, which was published in 1995. The report is based on a normative sample of more than 50,000 public safety applicants and supplements the CPI 434 with several advanced technical features designed to improving the accuracy and fairness of
the instrument in employment screening for public safety personnel. The features include risk estimates predicting the likelihood that an applicant will demonstrate particular counterproductive work behaviors, CPI scale profiles based on public safety normative data that can be compared to the scores of successful applicants, CPI scales based on demographic variables so that an applicants scores can be viewed controlling for sex and ethnicity, a list of CPI items endorsed by the applicant that are related to problematic work behavior according to scientific studies and expert psychologists, and lastly, a summary of CPI scales which indicate either favorable or unfavorable responses related to essential job functions and potential performance problems.

The CPI Police and Public Safety Selection Report was developed using preemployment test data of 50,488 applicants at over 100 public safety agencies for different public safety positions. Of the 50,488 applicants, 13,376 were hired and found to be successful. The majority of the sample, 40,814, was police officer applicants. Of the 40,814 police officer applicants, 10,680 were later determined successful incumbents.

The test responds directly to the central issues regarding test norms (e.g., representative, size, and specialized subgroups) according to Groth-Marnat (1997) in terms of selecting an appropriate test for a particular task (Roberts & Johnson, 2001). Additionally, in terms of clinical versus actuarial prediction, the CPI Police and Public Safety Selection Report uses formal prediction rules which, according to Groth-Marnat (1997) “can and should be used more extensively as a resource to improve the accuracy of clinical decision making.” These formal decision rules can be seen in the risk ratings, which according to Roberts and Johnson (2001) are “…generated from prediction equations that have been calibrated and cross-validated on large samples of previous applicants” (p, 2).
The CPI-based screening report is considered an essential part of the selection process and is conceptualized by the developers as a “professional-to-professional” consultation designed for the purposes of assisting in making hiring decisions. The test should not be used alone, but should be used in tandem with multiple additional data sources such as background information (e.g., personal and occupational history), a face-to-face structured interview addressing job relevant behaviors, and additional testing results (e.g., an objective psychological test intended for medical settings to assess for emotional stability or pathology). Decisions about employment suitability should be made based on the corroboration of several data sources. The central objective of the CPI special report is to assist psychologists involved in preemployment screening of police and public safety applicants as well as the psychological evaluation of incumbents for special assignments. The report is meant to aid in determining the “psychological suitability of the applicant for the position in question” (Roberts and Johnson, 2001, p. 4). The CPI is most helpful in assisting psychologists in preemployment suitability screening because the content of the scales are directly related to specific job requirements for police and public safety personnel.

The CPI Police and Public Safety Report consists of several unique features. The report features include the following: Identifying Information, which includes personal information such as name, social security number, age, gender, ethnicity, etc.; General CPI Results including the CPI Type and Level, selection relevant CPI items, and number of unanswered items; Job Suitability Snapshot, which “estimates various kinds of job-related risk associated with hiring the applicant” (Roberts & Johnson, 2001, p. 13); Applicant Comparison Profile # 1, which displays the applicants CPI scores graphically
in the form of two profiles relating the applicants scores to successful incumbent officers as well as members of the general community; Applicant Comparison Profile # 2, which again is reflected as two graphic profiles relating the applicants scores to successful incumbents in the same job category and incumbents of same gender and ethnicity for purposes of avoiding adverse impact; Applicant Level and Type Classification, which provides a graphic representation and brief summary description of both the applicant’s Level (v.3) and Type (v.1 and v.2) (e.g., Alpha, Beta, Gamma, Delta); CPI Special Purpose Scales such as Amicability, Hostility, Narcissism, Law enforcement Orientation, integrity, etc., which have particular relevance to organizational behavior, especially in policing; Selection-Relevant CPI Items which are items judged by a panel of psychologist experts as relevant to public safety work, answered in the “negative” direction by only ten percent of applicants, and correlated with substandard police performance: The item content of each atypical response and those correlated with substandard performance is displayed and organized into job function categories; Indicators of Essential Job Functions and Job Performance Problems for Police Officer Applicants presents scales shown to be significantly related to both police performance in terms of essential job function and specific behavior problems and lists them under two headings, Favorable indicators and Unfavorable indicators, as they are significantly related to satisfactory and unsatisfactory performance, respectively; Lastly, Item Responses lists all of the applicant’s response (true or false) to each of the 434 CPI items.

The specific focus of the present study is to examine the validity of the risk estimates that make up the Job Suitability Snapshot when applied to predicting the outcome (e.g., success or failure) of the background investigation. There are eight risk
estimates that compose the Job Suitability Snapshot. According to the authors of the CPI Police and Public Safety Selection Report, the first risk estimate has to do with an applicant’s likelihood of being rated “poorly suited” by a psychologist with expertise in police selection. There are six additional risk estimates that reflect the likelihood that an applicant has a history of “problem” behaviors thought to be relevant to the job of law enforcement officer. The estimates reflect the probability that an applicant will engaged in such problematic behaviors. The six problem behaviors are as follows: Job Performance, Integrity, Anger Management, Alcohol Use, Illegal Drug Use, and Substance Abuse. Lastly, for those individuals applying for law enforcement positions only, the CPI special report provides an estimate of the likelihood that an agency will eventually terminate the applicant if hired. The risk estimates are based on a prediction formula relating the CPI scales to specific outcome variables concerned with police and public safety work.

According to the authors of the CPI Police and Public Safety Selection Report Technical Manual (Roberts & Johnson, 2001), “Each risk rating is reported as a numerical probability of the undesirable outcome…,” which “are categorized into three risk level categories: High risk \((p > 50\%)\), Moderate risk \((p = 25\%-49\%)\) or Low risk \((p < 24\%)\)” (p. 15). The risk ratings were derived from prediction equations developed using logistic regression. The logistic regression equations were calibrated and cross-validated on the large number of police and public safety applicants that make up the normative sample for the CPI Police and Public Safety Report. In addition to providing the risk estimates for the particular applicant that is tested, the CPI special report also provides base rate values for the normative sample.
The sample used to develop the prediction formulas consisted of 50,488 police and public safety applicants. The poorly suited risk estimate was developed using a sub-sample of 22,867 cases and involved a completed psychological evaluation of each applicant that included a face-to-face structured interview, the CPI, and other tests such as the PAI, MMPI, STAXI, as well as a personal history questionnaire, all resulting in a suitability rating. The six additional risk estimates were developed using the Johnson-Roberts Personal History Questionnaire (PHQ) on a sub-sample of 36,276. The PHQ consists of approximately 300 questions addressing different job-related aspects of an applicant’s background, including education, employment, military experience, law enforcement experience, driving record, financial history, criminal record, substance use, etc. Each of the risk estimates reflect a composite variable derived from the PHQ that were identified as “problem” behaviors. The six composite variables reflect job-relevant problems in the following areas: Job Performance, Integrity, Anger Management, Alcohol Use, Illegal Drug Use, and Substance Abuse.

The risk estimates are considered valuable to the evaluating psychologist in that they reflect areas of concern in terms of behaviors that are considered problematic for purposes of law enforcement work and thus should prompt the evaluating psychologist to further investigate those risk estimates that fall in the “High” risk classification. Importantly, the psychologist using the CPI special report must be familiar with the behaviors that make up each risk rating and understand that each estimate is comprised of multiple behaviors, all varying in degree from seriously problematic to the more innocuous. It is the psychologist who makes the ultimate determination about whether or
not the behavior or combination of behaviors will be problematic for the job of police officer. In this sense, the actuarial technology is tempered by clinical judgment.

The previous several paragraphs have been devoted to a review of the CPI Police and Public Safety Report with an emphasis on the CPI risk estimates, which are the focus of this paper. As indicated above, the risk estimates of the Job Suitability Snapshot provide important information regarding various kinds job-related risks associated with hiring an applicant. Although the authors of the CPI special report have conducted several studies validating the risk measures, there remains a paucity of published research on the risk estimates. In the following section, the existing published and limited unpublished research on the CPI is reviewed. The lack of research on the CPI special report, specifically in terms of the suitability risk estimates, is reason for additional scientific inquiry.

The California Psychological Inventory and Police Selection

There is about a half of a century of professional recommendations and guidelines regarding the use of personality testing in police screening. Such extensive efforts by government officials as well as national and local accrediting bodies to oversee personnel selection procedures and to emphasize that personality assessment be included in such employment endeavors reflects the importance of personality testing in the evaluation of employment suitability of police officer applicants. There is also overwhelming empirical evidence to date supporting the use of personality testing in the selection of law enforcement officers (Verela et al., 2004). The CPI has been used in many studies as a predictor of police selection and of performance outcomes (James et al., 1984). The
instrument has been found time and time again to be a successful discriminator of suitable and unsuitable candidates. However, as indicated previously, only a small portion of this work has been published in professional sources (e.g., technical manuals, professional reference texts, science journals, etc). Nevertheless, a smaller, but still substantial body of published work exists supporting the use of the CPI in law enforcement screening. According to Roberts and Johnson (2001), “There is an extensive research literature supporting the validity and selection utility of the CPI scales…” (p. 21)

There have been many criteria used to investigate the validity of selection techniques. The two most frequently used criteria are supervisor ratings and evaluations during officer training (Gowan & Gatewood, 1995). Other criteria that have been used include civil service examinations, tenure, departmental awards, tardiness, absenteeism, turnover, number of arrests resulting in conviction, number of complaints, and disciplinary actions. Some authors contend that the many different outcome variables used in the employment selection literature represents a weakness that functions to limit generalizability and interferes with the comparisons of studies. However, hiring decisions are rarely made based on one criterion such as supervisor ratings or tardiness.

To date, no single behavioral indicator has been identified as the critical differentiator of suitable and unsuitable police officers (McDonough & Monahan, 1975). According to Campbell, McHenery, and Wise (1990), different personality traits are differentially relevant depending on the criterion used for job performance. Although certain behaviors may have more weight than others in terms of hiring, most selection decisions are based on a combination of behavioral indicators. Research employing narrowly defined criterion may misrepresent the usefulness of personality variables in
employment selection (Black, 2000). More comprehensive job performance criteria that include the many different facets of job performance would result in more accurate information in terms of the usefulness of personality measures in employment selection (Nathan & Alexander, 1988).

Conceptualizing counterproductive work behavior as a composite variable consisting of various behaviors thought to be related to poor work performance and problems in the work place is a more realistic approach. Looking at the research on counterproductive work behavior in aggregate allows researchers to view the literature in terms of the big picture, rather than focus of whether or not a particular behavior is an appropriate indicator, or whether the results of one study can be compared to the results of another based on the particular outcome variables. Therefore, the review of individual studies concerning the CPI and police work that follows will be summarized at the end of this section and the findings will be presented in composite form. The literature review will not be divided in terms of the different criterion variables used in the various studies. Instead, the research will be reviewed chronologically, starting with the earliest studies and progressing to the more recent.

*Study Descriptions and Findings*

Hogan (1971) noted that the first substantive work on the topic of personality testing in law enforcement settings was done by Matarazzo, Allen, Saslow, and Wiens (1964). Matarazzo et al (1964) conducted a study examining the relationship of several psychological tests (e.g., Wechsler Adults Intelligence Scale, Minnesota Multiphasic Personality Inventory, Edwards Personal Preference Schedule, Strong Vocational Interest Blank, and Rorschach) and successful police applicants. The authors found that police
officers are well adjusted in terms of their personality and above average intellectually. According to the authors, successful police officers’ are described by others as “active, conscientious, uncomplicated, and dominant (Matarazzo, 1964). The results of Matarazzo et al. (1964) challenged the then popular conception of police officers as uneducated, lower class, conservative, authoritarians. Although tests of intelligence such as the WAIS and projective measures of personality such as the Rorschach are not typically recommended or employed by law enforcement accrediting bodies or agencies, work like that of Matarazzo et al. (1964) laid the foundation for future research exploring the relationship between personality variables and police performance. Following Motarazzo et al. (1964), there were many empirical studies similar in nature. Several of those studies have examined the CPI in terms of predicting police officer work behavior (e.g., performance). Such findings are especially relevant to the preemployment psychological evaluation of police applicants.

Hogan (1971) conducted one of the first studies to examine the relationship between the CPI and performance of law enforcement personnel. The subjects consisted of three classes of police cadets at the Maryland State Police Academy (N = 141) and State Police with one year of experience (N = 42). The author compared cadet’s and officer’s CPI profiles to training staff and supervisor ratings, respectively. Product-moment correlations were conducted between the CPI standard scales and performance ratings. The largest correlations were found between CPI profiles and officer performance, which represents actual field work.

Eight scales (Wb, Re, Sc, Gi, Ac, Ai, Ie, and Py) were significantly correlated with officer ratings. The most significant correlation was that of the Intellectual
Efficiency (Ie) scale. Individuals scoring high on Ie tend to be described as “capable, confident, foresighted, and reasonable.” (Hogan, 1971, p. 683) When the author looked specifically at the relationship between cadet CPI scores and training performance, moderate but significant correlations were found between six scales (Do, Sy, Sa, Wb, Ai, and Ie) and training ratings. Lastly, three variables (Wb, Ai, and Ie) were found to significantly correlate with both cadet and officer performance ratings. The author concluded that highly rated police officers “have a sense of energy and good health (Well-being), are rational and independent (Achievement via Independence), and functionally intelligent (Intellectual Efficiency)” (Hogan, 1971, p. 684).

In additional to the primary analyses, the author developed a regression equation using four scales: Sp, Sa, Ai, and Ie. The regression equation, referred to now as the Police Performance Effectiveness Index (PPE), yielded a correlation of .42 (p<.01) with supervisors’ rating of effectiveness. The author suggested that the equation reflects an officer who is self-assured, pragmatic, interpersonally reserved, and motivated toward achievement.

The results of Hogan (1971) are supported by the previous findings of Matarazzo et al. (1964). Hogan (1971) also pointed out that the results of his study are consistent with the only other study of this kind at that time. The other study, conducted by Baehr, Furcon, and Froemel (1968), examined the validity of certain personality variables in predicting field performance. Baer et al. (1968) examined the personality characteristics of 512 Chicago patrolmen and compared their profiles with supervisors’ ratings. The authors concluded that Chicago patrolmen were stable, well-socialized, and family-oriented.
Hogan and Kurtines (1975) contended that their research extended previous finding (Hogan, 1971) regarding the personality variables of successful officers on the East coast of the United States to an organization of urban police officers on the West coast of the United States. Hogan et al. (1975) conducted a two-part analysis with separate samples aimed to address different questions. The sample consisted of 229 individuals. The first set of analyses consisted of 113 individuals (12 experienced officers, 31 cadets nearing the completion of academy training, 28 cadets in the beginning of training, and 42 applicants terminated early in the selection process) and examined personality variables related to “survival” of the selection process. In the second analyses, which consisted of 116 experienced officers, the authors examined personality variables related to effective performance.

The Hogan and Kurtines (1975) study was conducted with applicants and incumbents of the Oakland Police Department. The study was designed to address three issues. The first goal was to examine the modal profile of police officers. The second objective involved exploring the personality variables related to persistence in police work. The last target was to identify the characteristics correlated with effective performance.

The authors conducted a one-way analysis of variance to examine the difference between police officers and unsuccessful applicants. They found that unsuccessful applicants scored above the mean for men in general on the scales of Dominance, Sociability, Social Presence, Self-acceptance, Self-control, Achievement via Conformance, Achievement via Independence, and Psychological-mindedness. Conversely, unsuccessful applicant scored below the mean for men in general on the
Tolerance scale. Additionally, the two groups differed significantly on nine of the nineteen standard scales used. The unsuccessful applicant scored significantly lower on Dominance, Capacity for Status, Social Presence, Self-acceptance, Achievement via Independence, Intellectual Efficiency, Psychological-mindedness, Masculinity, and Empathy. These results indicate that, relative to successful applicants, the unsuccessful applicants were less assertive, had less potential for social mobility, were less socially poised and self-confident, had a markedly lower sense of self-worth, were less motivated toward individual achievement, were less pragmatic, less psychologically minded, were less masculine, and less insightful socially.

Next, the authors conducted a correlational analysis to explore the personality variables associated with effective performance. The strongest predictors of effective performance were the scales Capacity for Status, Achievement via Independence, and Intellectual Efficiency. The authors noted that the results from the current study are consistent with those found in a previous study conducted on Maryland police officers by the first author. Taken together, the two studies indicate consistency across states and departments.

Lastly, the authors looked at two CPI-based regression equations, one for estimating social maturity and the other was an index of leadership, for both the Maryland officer sample and the Oakland officer sample. The mean scores for both groups of officers on the social maturity estimate were lower, but not significantly lower, than the community average indicating that police officers are neither more nor less socially mature than the average individual. When examining officer scores on the leadership estimate, both Maryland and Oakland police officers scored significantly
above the community average on the leadership index. While the Maryland sample scores on the index were equal to the scores of the sample of leaders originally used to develop the index, the Oakland sample scored notably higher than Gough’s original leadership criterion group. The authors concluded that experienced police officers have “considerable force of character and leadership potential” (p. 293).

The Hogan and Kurtines (1975) study provides information about what type of individuals become police officers, what characteristics are related to qualified and unqualified applicants, and what personality factors are related to successful performance. The authors discovered that within their sample, police officers tended to be masculine, self-confident, and socially competent. Additionally, applicants who survived the selection process scored significantly higher on CPI variables such as assertiveness, social poise, self-confidence, motivated toward individual achievement, and intellectual efficiency than their nonqualified counterparts. Lastly, the characteristics of police officers who were successful in performing their job were “…functional intelligence, achievement motivation, and social poise.” (p. 289) Much of the research on personality testing in law enforcement setting is focused on officer performance in the academy and in the field. Personality variables linked to later performance are also considered valuable indicators in the selection of officers. The study conducted by Hogan and Kurtines (1975) provides support for the rationale that the characteristics considered important in the determination of an applicant’s suitability are the same that are found empirically related to future performance.

In an unpublished doctoral dissertation, Hortsman (1976) conducted a study administering the CPI to police officer recruits upon entry into the police academy and
comparing the results with performance evaluations upon their completion of the academy. The author found a correlation of .61 with 37% of the variance in performance evaluation explained by the CPI scales. The Well-being (Wb) scale was most highly correlated with performance evaluation. The findings of Hortsman (1976) support the previous findings by Hogan (1971) that the CPI Well-being scale is significantly related to cadet performance in academy training. As seen in both Hogan (1971) and Hogan and Kurtines (1975), there are several other variables that have been found significantly related to police officer performance, all of which should be considered when deciding which applicants will be selected for the challenging and critically important job of law enforcement officer.

Lietner and Sedlacek (1976) conducted a study examining the usefulness of seven different personality and attitude measures, one of which was the CPI, to predict various officer performance criteria. The criterion variables included tenure, commendations, reprimands, absenteeism, ratings by supervisor as closest to the "ideal" officer, promotions, peer and self-ratings, as well as supervisor ratings. The study sample consisted of 52 campus police officers at the University of Maryland. The authors used multiple regression analysis to examine the degree to which the several measures predicted the various officer performance criteria. The authors found that, in terms of the CPI, scores on the Dominance scale were significantly related to officer absenteeism. Lower scores on the Dominance scale were related to higher rates of absenteeism. The results of Lietner and Sedlacek (1976) support the findings by Hogan and Kurtines (1975) and Hogan (1971) that the Dominance scale of the CPI is an important variable in the prediction of successful police performance.
A study conducted by Mills and Bohannon (1980) provide additional support for the validity of several CPI scales in the prediction of police officer behavior. The authors claimed that past research suggests practical intelligence, self-assurance, good interpersonal skills, and achievement motivation are generally related to leadership and successful performance in police officers. Mills and Bohannon (1980) investigated the personality variables related to leadership and overall performance in a group of 49 male police officers from an East Coast department. The study was also designed to examine the predictive validity of both Gough’s (1969) leadership equation and Hogan’s (1971) Police Performance Equation. All subjects completed the CPI after beginning the police academy. After one year of service as a police officer, multiple supervisor ratings were obtained regarding leadership and overall suitability for each officer. Reliability coefficients for supervisor ratings were .78.

The authors found that the sample of police officers scored at least .5 standard deviations higher than the community mean on the Dominance and Self-control scales of the CPI. Leadership as measured by supervisor ratings was associated with the CPI scales of Tolerance, Achievement via Independence, and Intellectual Efficiency. The overall suitability ratings correlated .84 with ratings on leadership. Overall suitability ratings were related to higher scores on Socialization, Tolerance, Communality, Achievement via Independence, Intellectual Efficiency, and Flexibility. Officer’s scores on the Leadership Index were not significantly correlated with ratings of leadership or overall suitability. Thus, the Leadership Index accounts for an insignificant amount of the variance in supervisor ratings. Conversely, Hogan’s Police Performance Equation was significantly correlated to both ratings of leadership and overall suitability.
In summary, the Police Effectiveness Equation (Hogan, 1971) was significantly more predictive of leadership than the leadership index (Gough, 1969). According to Mills and Bohannon (1980), both Achievement via Independence and Intellectual Efficiency have been repeatedly linked to effective police performance. The authors further note that contrary to the popular belief that police are rigid, hyper masculine, thrill seeking, dictators, law enforcement officers of today are “bright, assertive, autonomous, self-assured, responsible, and level headed individuals.” (p. 683) Past research on personality variables and police performance have provided valuable information regarding the most appropriate selection tools for identifying individuals who are well suited for and likely to succeed in law enforcement work. In addition to the evidence that certain CPI variables are related to police officer performance, and the implication that personality instruments could be useful in the screening of law enforcement officers to identify those applicants who demonstrate characteristic features of problematic officers, there is research showing an associated between CPI variables and employment selection decisions. The evidence points to an agreement between hiring decisions based on a personnel selection interview panel and the CPI variables.

The POST psychological screening manual (Hargrave & Berner, 1984) was the result of a large project that included several components related to law enforcement research. One aspect of the study included predictive academy studies that related test and performance variables of cadets who had not been psychologically screened as a basis for selection. Another aspect of the study included an incumbent officer study, which examined predictive relationships between preemployment test scores and subsequent performance measures. Hargrave and Berner (1984) examined the
relationship between CPI scales and both academy attrition and leadership ratings in a sample of academy graduates from three different police academies. The authors found that academy graduates scored significantly higher than those candidates who attrited on CPI scales Sp, Ie, Sy, So, To, Cm, and LPI.

James, Campbell, and Lovegrove (1984) conducted a study investigating the relationship between interview decisions and scores on the CPI. The authors looked at 279 male and female applicants to the Victoria Police Department. The selection process for the Victoria Police Department included physical ability testing, educational testing, a medical examination, and an oral board interview. The CPI was given after the educational testing and not used in the determination of applicant’s suitability as the department did not employ psychological testing at that time. The applicants were informed that CPI results would have no bearing on the selection process. James et al. (1984) conducted two-tailed t-tests computing the CPI scale scores between those applicants that succeeded and those that failed the selection interview. The authors found that men who were successful in the interview scored significantly higher at the .01 level on the scales Wb, So, and Sc. Successful men also had significantly higher scores at the .05 level on CPI scales Re, To, Gi, Ac, and Social Maturity. The strongest positive associations with interview ratings were found within Gough’s Class II scales, which represent responsibility and socialization. The authors noted that no difference was found between successful and failed women.

The James, Campbell, and Lovegrove (1984) study provides support for the validity of the CPI in preemployment selection of law enforcement officers. However, one important issue to consider when looking at police performance research is the
possible change in predictors of successful performance over time (Pugh, 1985). It may be the case that different CPI scales, indices, or algorithms (e.g., risk ratings) predict different facets of performance at different times. One example of the change in predictors over the course of an officer’s service can be seen in a study conducted by Pugh (1985).

Pugh (1985) looked at the CPI in regard to its ability to predict job performance at different times over the course of an officer’s career. Subjects were selected from a large applicant pool of two police recruit classes from the Edmonton Police Department in Alberta Canada. After meeting certain prerequisites, 61 applicants were included in the study. Subjects were rank-ordered and placed in one of three performance groups (high performers, average performers, and low performers) by the sergeants, senior constables, and staff sergeants. Officer performance was judged at 2 and 4.5 years. Step-wise discriminant function analysis was used to identify differences among the three criterion groups and select the most predictive personality variables for the three levels of job performance.

Pugh (1985) found support for the hypothesis that the specific personality features required for high performance are different at different times in an officer’s career. After two years on the job, higher scores on Capacity for Status (Cs) was the best predictor of job performance. The results of Pugh (1985) indicate that those officers with the ability to strive for status are rated higher on job performance. The author suggested that police officers who are most focused on fitting in (e.g., being trusted member of the team) are most successful during the first couple of years on the job. After four and a half years on the job, both the Responsibility (Re) and Well-being (Wb) scales significantly
differentiated among the three performance groups. Also, the Socialization scale approached significance. The best predictor of high performance after four years on the job is a stable, responsible, and socially skilled individual (Well-being, Responsibility, and Socialization). Therefore, the mature and responsible employee who is focused primarily on the job is the highest performer after the initial career stage.

Although the Pugh (1985) study explains that the predictors indicating high performance change over time, it is apparent that certain variables are important in surviving the probationary period and thriving over a lifetime of police work. It is those variables that predict future adjustment and job success that should be used in determining an individual’s suitability for police work. Therefore, characteristics such as Capacity for Status, Responsibility, Wellbeing, and Socialization should be considered valuable qualities for prospective law enforcement officers. Those same qualities should be the targets for screening and subsequent hiring decisions.

In yet another study, Hargrave, Hiatt, and Gaffney (1986) compared MMPI and CPI test profiles of state traffic officers to those of deputy sheriffs; two groups representing very different law enforcement activities. The subjects consisted of 691 cadets. None of the officers had undergone psychological screening as a condition of employment. However, all officers had completed the agencies preemployment screening, which included an oral board review, reading and writing tests, physical ability assessment, medical examination, and a thorough background investigation. Additionally, both groups had completed academy training.

The authors looked at both mean profile differences as well as profile differences of officers who were rated high and low by their primary academy training instructors.
Ratings consisted of a 5-point behaviorally-anchored scale assessing emotional suitability. The MMPI and CPI scale scores were used as well as several supplemental scales. Discriminant function analyses were performed to differentiate officer groups and analyses of variance were conducted comparing officer personality test scores from both the high and low rated groups. The results of the discriminant function analyses were significant. The discriminant function calculated for the MMPI correctly classified 65 percent of the subjects into the two groups. The discriminant function for the CPI correctly classified 62 percent of the subjects. The 11 CPI scales that contributed significantly to the classification include Do, Sy, Sp, Sa, Wb, Ac, Ai, Fx, In, Mi, Lead.

Comparing officers rated high and low on psychological suitability by training instructors provides meaningful information regarding the personality variables related to effective and ineffective officers. For the CPI, those officers rated high on suitability tended to score higher on most scales. Highly rated officers scored significantly higher than their low rated counterparts on Ac (Achievement via Conformance), Ai (Achievement via Independence), Wb (Wellbeing), Mi (Managerial Interests), and Lead (Leadership). Therefore, according to Hargrave et al. (1986), characteristics such as achievement orientation, work ethic, organization, ambition, and leadership potential are indicative of highly rated officers despite group affiliation.

The results of Hargrave et al. (1986) demonstrate yet again that successful officers are achievement oriented. Such officers are interested and thus driven to be successful by agency or self standards. Both those who achieve by way of conformance to organizational standards and those who strive to be independently successful are equally highly suitable according to Hargrave et al. (1986). Another finding consistent
with previously presented studies is that high performing police officers, as evidenced by supervisor ratings, have a higher personal sense of wellbeing than their lower performing counterparts. Such officers are comfortable and generally satisfied with themselves and their lot in life. They have few doubts or worries and tend to have high esteem and confidence (e.g., spirit/morale) (McAllister, 1996). One notable finding across several studies is that applicants retained for employment and considered suitable trainees and high performing officers score higher on most all scales compared to their unsuitable counterparts. Additionally, there are several studies providing positive evidence that various other scales from the CPI are useful in determining the suitability and later success of police officer applicants.

Fitzgerald (1986) examined the relationship between selected scales on the MMPI and CPI and aspects of police officer performance. The study sample consisted of 90 police officers from 6 municipal police departments in St. Louis, Missouri. The personality measures were administered prior to employment. Stepwise multiple regression analyses were conducted. The author found that lower scores on the CPI Responsibility scale, which is considered a measure of maturity, were significantly correlated with more citizen complaints, more disciplinary actions, and higher rates of absenteeism compared to fellow officers. In addition to the Responsibility scale, other variable were related to performance problems. Lower scores on CPI scales Self-control (Sc), Dominance (Do), and higher scores on the Good Impression (Gi) scale were related to a higher rate of annual questionable sick days. Officers with higher scores on CPI scales Capacity for Status (Cs) and lower scores on Communality (Cm) received more disciplinary actions than other officers. The authors concluded that “some personality
scales, particularly those of the CPI, do have moderate correlations to aspects of problem performance, with the CPI Responsibility scale emerging as the strongest scale." (p. 2).

Another study reference in an unpublished doctoral dissertation by Schneider (2002) is Geraghty (1986). Geraghty (1986) conducted a study on a sample of 140 officers looking at the CPI as a predictor of police officer performance. The author found that officers scoring higher on the CPI Responsibility (Re) scale were rated higher by superiors than their lower scoring counterparts. Schneider (2002) noted that the findings “reiterated the value of officers being high on …Contientiousness …with regard to supervisory ratings.” (p. 36) The results are consistent with Fitzgerald (1986) and several other previously cited research indicating that the Responsibility scale of the CPI is a significant predictor of police performance.

Research investigating performance predictions based on test profiles (Hargrave 1985; Hargrave & Berner 1984) and both test and interview data combined (Hiatt & Hargrave, in submission; Roberts, 1985) have yielded significant results for officers in training and those on the job. Hargrave and Hiatt (1987) conducted a study investigating the relative contributions of a semi-structured interview (supplemented with a life history questionnaire), the MMPI, and the CPI in predicting performance in two classes of academy cadets.

The Hargrave and Hiatt (1987) study consisted of 95 subjects. None of the cadets had undergone psychological screening as a condition of employment, but all had undergone agency employment selection procedures. All cadets were tested and interviewed by two psychologists at the beginning of training. Each candidate was given a rating of either suitable or unsuitable for law enforcement work. Suitability
determinations were based upon results from tests or interview information only, then upon a combination of the two data sources. All cadets were given a training rating of successful or unsuccessful. Training ratings were based on training attrition, training officer’s rating on a 5-point behavioral scale, and peer evaluations. The suitability ratings were compared to academy training ratings of successful or unsuccessful. Additionally, analyses were conducted looking at those applicants who would have been disqualified by psychological screening, but who were rated as successful performers.

For the CPI measures, the successful subjects all scored higher than the unsuccessful subjects. Three of the standard CPI scales differed significantly between the successful and unsuccessful groups. The scales include Sp, Cm, and Ie. In addition, two supplemental scales showed significant differences between groups. These were Wo (Work orientation) and Lead (Leadership Index). In the case of false positives (e.g., successful subjects who were rated unsuitable), six subjects had an average of three CPI scales below a standard score of 40; the most frequent low scales for the false positive group were Sy and Ie.

The results of the Hargrave and Hiatt (1987) study suggest that successful officers, as determined by the completion of academy training, supervisor ratings, and peer ratings, scored higher than their unsuccessful counterparts on several CPI scales. Successful police trainees scored significantly higher than unsuccessful trainees on CPI scales measuring Social Presence, reflecting poise, spontaneity, and self-confidence in their dealings with others (McAllister, 1996). Successful trainees scored significantly higher on the Communility (CM) scale, which reflects the modal profile for the normative sample (e.g., being similar to most others). Successful police trainees also
scored significantly higher on the CPI scale Intellectual efficiency (Ie), so named for the scales measurement of intellectual orientation and the efficient employment of cognitive resources. Such individuals are well organized and successfully strategic in their resourceful approach to problem solving. The elevated special purpose scales of Wo and Lead indicate that successful trainees were hard working, disciplined, reliable, and perform well. They were also energetic, confident, assertive, optimistic, and able to gain the cooperation of others. Lastly, and of particular relevance to the current study, was the existence of false positives that shared two significantly low CPI scales: The Sociability (Sy) and Intellectual Efficiency (Ie) scales. Therefore, those applicants that were successful, but were rated unsuitable tended to be private, socially reserved, loner types who are uncomfortable around strangers and in crowds. The false positive subjects scored in a manner reflecting less organized and inefficient individuals who are haste and tend to think in black or white (McAllister, 1996).

Hiatt and Hargrave (1988) examined the job performance of 55 police officers in an urban law enforcement agency. The sample consisted of two groups: 15 incumbent officers who were hired despite being judged unsuitable by the evaluating psychologist and 40 officers hired after being judged suitable. All officers had worked long enough to receive at least one performance evaluation. Officers were rated as satisfactory or unsatisfactory based on personnel files. Ratings were made without knowledge of the psychologist’s judgment. A rating of satisfactory was given to those officers who had never received a disciplinary action and had no more than one rating of below satisfactory on any performance evaluation. A rating of unsatisfactory was given to officers who received any of the following: 1) multiple below satisfactory ratings on
performance evaluations, 2) suspended or asked to resign in lieu of termination, or 3) convicted of off-duty law violations.

Thirty-one officers were given a rating of satisfactory and twenty-four were rated unsatisfactory. The mean scale scores on the MMPI and CPI were compared for both satisfactory and unsatisfactory job performance groups. The authors reported finding statistically significant difference between the two performance groups. The authors reported finding significantly lower scores on the Ai (Achievement via Independence) scale for the unsuitable performance group. The significantly higher score on Ai indicates that the satisfactory officers were more mature, forceful, independent, and self-reliant (Gough, 1975). Additionally, satisfactory officers scored higher on 13 out of 18 CPI scales. The unsatisfactory group scored lower on all scales except Do, Cs, Sa, Gi, and Cm. However, these differences did not reach a level of significance. Although the sample size was small in this study, the results support past findings that Ai can differentiate high performing officers from their low performing counterparts (Hogan, 1971; Hogan and Kurtines, 1975; Mills and Bohannon, 1980) providing further support for the use of psychological tests such as the CPI in the psychological screening of law enforcement officers.

Hargrave and Hiatt (1989) published a study that consisted of two separate investigations of the CPI in law enforcement officer selection. The first study consisted of 579 cadets from three different law enforcement academies. None of the agencies used psychological tests in the screening process. All subjects were administered the CPI on the first day of training. At the end of their training, each subject was rated in regards to their psychological suitability by their primary training instructor. The ratings of either
not suited or suited were used as outcome criteria. A multivariate analysis of variance (MANOVA) was conducted comparing CPI scores of those found psychologically suited to those found psychologically unsuited. Additionally, peer ratings were collected as a measure of interrater reliability.

The authors reported that 13 percent of the subjects were rated as psychologically unsuited for law enforcement work. The results of MANOVA indicate that nine scales were found to significantly differentiate the two groups at the .05 level of statistical significance. The scales were Sy, Sp, Wb, Cm, Ac, Ai, In, Mp, Wo. Additionally, four of those scales (Sp, Cm, Ac, Ie) differentiated between groups at the .01 level of significance. The authors also found that with the exception of four scales (Fe, Fx, Gi, Sc), all mean scales scores were higher for the suited group. Lastly, both Gough’s LPI scores and Hogan’s PPE scores significantly differentiated between the two groups; higher rated individuals scored higher on both indexes.

The authors concluded that law enforcement officers found psychologically unsuited for duty scored lower on most CPI scales than those rated as psychologically suited. The authors found that this relationship held for all of Gough’s clusters, with the exception of Class IV, as well as the supplemental scales and indexes. Officers who demonstrate qualities such as self-confidence, poise, maturity, self-control, personal values, and achievement potential were considered more psychologically suited by training instructors for law enforcement work than those lacking such qualities. The authors noted that the group differences on Sy, Sp, Wb, Cm, and Ie found in this study was consistent with findings from past research in training settings (Hargrave & Berner, 1984; Hogan, 1971). The results from study 1 demonstrate that “… social confidence,
independence, and poise; seeing oneself as similar to others; and comfort and persistence in structured settings.” (p. 275) are characteristic features differentiating the two groups. Additionally, the LPI, which is designed to assess leadership, foresight, and decision making ability, reflecting attributes such as “dominance, self-confidence, and ability to think clearly…” (p. 275) was significantly different for the two groups. Lastly, study 1 produced results similar to those of Hogan (1971), Hogan and Kurtines (1975), and Mills and Bohannon (1980), finding PPE to be significantly higher for officers rated suitable than for those rated unsuitable.

The second investigation undertaken by Hargrave and Hiatt (1989) consisted of 45 incumbent officers from three different municipal law enforcement agencies. The authors investigated the CPI profiles of officers who had displayed seriously problematic work behavior and compared their profiles to a matched sample of non-problematic officers. The subjects were matched by agency, date of hire, sex, age, race/ethnicity, and education. All officers were subject to extensive screening procedures including an oral board interview, physical abilities test, medical examination, test of reading and writing skills, and a full-field background investigation. The subjects were also administered the CPI as part of the screening process. The problematic officers were those who exhibited such counterproductive work behavior as illicit relationships with prisoners, drug smuggling, illegal drug use, unnecessary use of force, physical altercation with other officers, and violations of agency regulations that resulted in prisoner escape. The results of problematic behavior lead to termination, resignation in lieu of termination, and suspension without pay.
After comparing the mean scores for the different classes for each group, the authors found that Gough’s Class II variable were found to have the most predictability. Next, a MANOVA was conducted to compare the Class II scales for the two groups. The authors found that non-problematic officers scored higher on all Class II scales except for Cm. Significant group differences were found on scales Wb, So, and Sc. Scale To was also found to near significance. The two groups differed substantially in that non-problematic officers appear much more comfortable with themselves and their current life situation; their moral is better and they are more tolerant, trusting, and diplomatic. Conversely, problematic officers are less dependable and less likely to adhere to social norms. Further, problematic officers are less disciplined, stable, and deliberate. In accordance with previous research (Fitzgerald, 1986; Hogan, 1971; Hortsman, 1976) the study found Sc to be of particular importance. The CPI scale So was also found to be strongly related to job performance; A finding consistent with past research as well. Together, Sc and So assess self-regulation/control, level of impulsivity, risk taking/thrill seeking, selfishness, objectivity, honesty and dependability, adherence to norms/rules, and manipulation/opportunism (McAllister, 1996).

One study, unique to the published literature on the relationship between the CPI and police performance, was conducted by Wright, Doerner, and Speir (1990). Wright et al. (1990) investigated the relationship between preemployment MMPI and CPI scores and BARS scores of police recruits obtained during field training officer (FTO) program. In the Wright et al. (1990) study, the authors found no relationship between CPI scales and behaviorally anchored ratings of police trainees.
The study was conducted at the Tallahassee Police Department, in Tallahassee, Florida. The FTO program was modeled after the same program used by the San Jose Police Department and other municipal police agencies. All officers completed the MMPI and CPI as one part of the screening process. Those applicants found unsuitable for police work were excluded from the study. The criterion measures were obtained from personnel files and consisted of FTO daily training evaluations of 33 items within five behaviorally anchored performance dimensions. The dimensions include appearance, attitude towards criticism and police work, knowledge of policy and law, field performance, and interpersonal relationships with citizens and fellow officers. Scores on each item range from one to seven. Scores of four reflect average performance whereas scores near seven indicate superior performance. Consistent ratings below four result in termination.

The authors conducted a series of correlational analyses and found virtually no relationship between either of the personality measures and the BARS measure of field training performance. The authors concluded that psychological test scores cannot be used to make accurate predictions about rookie performance in training. In light of the extensive research to the contrary, the results elicit further examination. There are several possible hypotheses for the disparate results. Wright et al (1990) proposed one explanation for the study results, positing that the passage of time between when the psychological test results were gathered and when the performance evaluations were made, may have affected the test score reliability. The authors speculated that the socialization that occurs during induction into the police subculture may be a more powerful factor in terms of thinking, feeling, and behaving than we currently understand.
Another hypothesis, and a common problem in similar research designs, is that the majority of applicants who would have been unfit for police work if hired, were screened out in the initial preemployment phase. The issue of range restrictions was discussed previously and will continue to be addressed throughout the present study.

A study conducted by Wells (1991) looked at both the MMPI and the CPI in an attempt to clarify previous research findings regarding the measures as predictors of police performance. Data was sampled from seven different police departments. The predictors were scores on the MMPI and CPI, which were administered prior to being hired. The criterion consisted of officer performance classification of either problem or conventional. The classifications were derived from negative on the job behaviors (e.g., oral reprimands, written reprimands, suspensions, citizen complaints, substantiated citizen complaints, motor vehicle accidents, and termination) defined by police supervisors.

The author conducted discriminant function analyses and found that overall performance classification (e.g., problem or conventional) was not predicted by CPI and/or MMPI scores. However, whether or not an officer had engaged in any of the problem behaviors was significantly predicted by scores on the personality measures. The CPI scale Ac (Achievement via Conformance) was a significant predictor of substantiated citizens complaints. The author found that officers scoring lower on Ac were more likely to have at least one substantiated citizen complaint in their record. Additionally, The CPI scale of Cm (Communality) was a significant predictor of citizen complaints. Wells (1991) reported that the CPI scales may have a curvilinear relationship with police performance. The author concluded that although the MMPI may be useful for detecting
obvious pathology, the CPI scales seem to be better at measuring those personality traits most important for the successful performance of the job of law enforcement.

In a more recent study, Sarchione, Cuttler, Mucinsky, and Nelson-Gray (1998) examined the validity of personality and life history construct in predicting counterproductive work behavior in police officers. The authors hypothesize that the construct of conscientiousness as measured by the three CPI scales and three life history dimensions, would discriminate between those law enforcement officers who engage in counterproductive behavior and those who do not.

The study consisted of two criterion groups, disciplinary (n = 109) and control (n = 109). The disciplinary group consisted of officers formally disciplined for behaviors such as sexual misconduct, substance abuse, insubordination, embezzlement of property, truthfulness, multiple motor vehicle violations, inappropriate verbal conduct toward the public, and multiple duty violations. The control group included active duty officers who did not exhibit the dysfunctional job behaviors identified above. The predictors consisted of the three scales on the CPI 434 Form (e.g., Responsibility, Socializations, and Self-Control) and life history information obtained from a personal history questionnaire, structured interview, and background investigation. Life history information was organized into three rationally derived domains (e.g., work history, criminal history, and drug use history) by subject matter experts.

The authors computed correlations between the six predictors and conducted univariate analysis (t-test and effect sizes) to assess the capacity for the predictors to differentiate the criterion groups. The results show that the three CPI scales significantly differentiated the two groups. The Responsibility scale was the strongest predictor. The
three life history indices also significantly differentiated the two groups. All results were consistent with the author’s hypotheses. The results from this study suggest that conscientiousness is predictive of dysfunctional job behavior in a law enforcement sample. In terms of CPI scales, individuals who are careless, impulsive, and have little sense of duty (Re) are most likely to evidence counterproductive work behavior. Those officer applicants who are more rebellious and prone to take risks (Socialization: So) as well as those who are unpredictable and excitement-seeking (Self-Control: Sc) are also more likely to engage in dysfunctional work behavior. In terms of the construct-oriented life history indices, past behavior predicted future behavior in that those individuals evidencing past problems with employment, crime, and drugs were problematic as police officers from a disciplinary perspective.

Schneider (2002) conducted a study examining the degree to which the Big-Five personality factors, as represented by the MMPI, CPI, and IPI, predicted a variety of police officer job performance criteria. The predictors consisted of the MMPI, CPI, and IPI scales. The personality inventory scales were grouped in terms of the five factors that make up the Big-Five personality taxonomy. Both the scores on the Wonderlic Personnel Test and an overall fitness rating rendered by a psychologist examiner were used to assess the variance accounted for by cognitive ability and the clinical interview, respectively. The criterion for the study consisted of civil service examination scores, academy grades, supervisory ratings of overall job performance, and termination. Archival data were utilized. The sample consisted of 270 incumbent police officers from a large Southeastern United States municipality. After a comprehensive literature review, the author hypothesized that individuals scoring higher on Extroversion,
Conscientiousness, Agreeableness, Openness to Experience, and lower on Neuroticism, would perform better across job performance criterion than their peers scoring in the opposite direction on the personality inventory scores representing the Big-Five factors. The literature review was directly related to the CPI and police candidate and incumbent officer performance.

The author created a classification of scales along the Big-Five factors via content analysis. Extroversion was represented by the CPI scales of Capacity for Status (Cs), Sociability (Sy), and Social Presence (Sp). Neuroticism was represented by the CPI scales of Self-acceptance, Independence, and Well-being. Contentiousness was represented by the CPI scales of Responsibility (Re), Socialization (So), Self-Control (Sc), Communality (Cm), and Achievement via Conformance (Ac). Agreeableness was represented by the CPI scales of Empathy, Tolerance, Good Impression, and Flexibility. Lastly, Openness to Experience was represented by the CPI scales of Achievement via Independence, Intellectual Efficiency, Psychological-mindedness, and Femininity. The author conducted a series of hierarchical regression analyses. A separate analysis for each personality inventory was conducted for each performance criteria to determine which of the three inventories best predicted performance.

In general, the author found that the Big-Five factors of Neuroticism (e.g., emotional stability), Conscientiousness, Agreeableness, and Openness to Experience were significant predictors of several police officer performance criteria. The CPI was the strongest predictor of supervisory ratings. Additionally, the CPI was a better predictor than the MMPI on all other performance criteria. However, the CPI did not demonstrate significant correlations with the remaining criterion. Lastly, the CPI did not out predict
the IPI on any criterion other than supervisory ratings. It appears that the contentiousness personality construct as represented by the CPI scales of Responsibility, Socialization, Self-Control, Communality, and Achievement via Conformance, was a significant predictor of supervisory ratings.

The Schneider (2002) study provides support for the use of several Class I scales (e.g., Responsibility, Socialization, Self-Control, Communality) as well as the Class III scale Achievement via Conformance. The results are consistent with Sarchione et al. (1998) in that the CPI scale that appear to reflect the contentiousness construct are valid predictors of police officer performance. The results are also consistent with the findings from Cuttler and Muchinsky (2006) that three selected CPI scales (e.g., Responsibility, Socialization, and Self-Control) representing the contentiousness construct were significant predictors of officer misconduct on the job.

Surrette and Serafino (2003) conducted a validity study looking at the relationship between personality measures and tests of cognitive ability and police officer performance after one year on the job. The study sample consisted of 129 police officers from a variety of small law enforcement agencies in New Mexico. The police officer applicants were hired in the mid to late 1980s. The predictors were three personality measures and two cognitive ability measures. The personality measures used included the Minnesota Multiphasic Personality Inventory, the California Personality Inventory, and the Inwald Personality Inventory. The two measures of cognitive ability used included the Shipley Institute for living Scale, which is a measure of general cognitive ability and the Nelson Denny, which is a test of reading ability. The criterion variable consisted of supervisor ratings on the overall performance of each officer after one year of
employment. Based on the results the authors claimed that neither personality nor cognitive ability was significantly related to on-the-job performance of police officers after one year of employment. In terms of both cognitive and personality variables, the results are inconsistent with the majority of past research.

Ones, Viswesvaran, Cullen, Drees, and Langkamp (2003) conducted a meta-analysis on the validity of personality variables in the prediction of police officer behavior. The predictors used included the MMPI and the CPI. The criterion used was productive and counterproductive work behavior. Counterproductive work behavior consisted of behaviors such as misuse of firearms, inappropriate sexual behavior, integrity problems, insubordinations, and violation of department regulations. The authors found a negative relationship between counterproductive work behavior and the personality variables agreeableness, impulse control, and socialization. The authors also found a positive relationship between risk taking and counterproductive work behavior. The results from Ones et al. (2003) indicate that personality variables are useful in the prediction of a wide variety of behaviors linked to problems in the effective performance of police work.

The most comprehensive review of the literature to date regarding law enforcement selection was conducted by Michael Aamodt (2004). Aamodt (2004) conducted a meta-analysis of the research looking at the CPI in police selection. Meta-analysis is a statistical method for combining research results across a large number of studies: It provides a useful means for understanding the research findings in the area of police selection. The author concluded that several scales on the CPI are significantly related to supervisor ratings of performance, academy performance, or disciplinary
problems. The author reported that the most notable scales appear to be Tolerance and Intellectual Efficiency. Both Tolerance and Intellectual Efficiency are significantly correlated with supervisor ratings, academy grades, and discipline problems. According to Aamodt (2004) “people scoring high in tolerance are tolerant, non-judgmental, and resourceful and those scoring high in intellectual efficiency are intelligent, clear thinking, and capable.” (p. 100).

Enright (2004) collected personality test scores from 218 police officers at two different law enforcement agencies. The authors then examined the relationship between pre-hire personality test scores and subsequent on-the-job performance. The personality tests used were the MMPI and the CPI. On-the-job performance took the form of either positive (e.g., motor vehicle accidents, written reprimands, suspensions from duty, and terminations/firings) or negative (e.g., written commendations, major commendations, or promotions) officer performance. The author used Structural Equation Modeling to evaluate model fit and strength of relationship between the predictors and criterion. The proposed model was inadequate. The authors did not find a significant relationship between the personality variables thought to represent a prosocial construct and a construct representing psychological distress and positive and negative officer performance variables.

Kostman (2004) conducted a study examining the usefulness of the “Job Suitability Snapshot” on the PAI and CPI special police and public safety reports in predicting officer’s fitness for duty. Although the Kostman (2004) study was an investigation of the psychological stability of incumbents rather than suitability of applicants, it is a sound assumption that the reason or reasons for which an individual
would be found unfit for police work would represent disqualifying criteria for those applying for law enforcement work. The author collected data from a psychological assessment center in Chicago, Illinois providing preemployment and fitness-for-duty psychological evaluations for small and large law enforcement agencies. The study consisted of 138 randomly selected police officers from a large metropolitan police department in the Midwestern United States. All participants had been ordered to undergo a fitness-for-duty evaluation between 2002 and 2004. The sample consisted of 82 officers classified as fit and 55 officers classified as unfit.

The author looked at the correlation between subject variables (e.g., fitness status, gender, age, ethnicity, years on the force, level of education, and reason for referral) and seven PAI and CPI risk estimates from the “Job Suitability Snapshot” (e.g., Probability of being rated a “poorly suited” applicant by psychologists with expertise in public safety screening, Job performance problems, Integrity problems, Anger management problems, Alcohol use concerns, Illegal drug use concerns, and Substance abuse proclivity). Logistic regression analyses were conducted with fitness status as the criterion. The author found that “CPI scores accurately predicted 67.4 percent of fitness categories…” (p. 89). The CPI risk estimates for CPI 1 (Probability of being rated a “poorly suited” applicant by psychologists with expertise in public safety screening) and CPI 3 (Integrity problems) were significant predictors. The results indicate that for each unit increase in CPI 1, the officer was 1.08 times more likely to be unfit, and for each unit increase in CPI 3, the officer was 1.06 times more likely to be fit. The results of the Kostman (2004) study indicate that both the PAI and CPI “job suitability” risk estimates do an equal job of predicting officer’s fitness for duty.
Cuttler and Muchinsky (2006) conducted two studies looking at the prediction of police performance. The first study investigated the predictability of law enforcement candidate’s success or failure of academy or field training based on personality, mental ability, and life history variables. The second study examined the predictability of disciplinary problems in police officers using personality, mental ability, and life history variables.

The first study was designed to differentiate between those applicants who failed versus those who passed the law enforcement academy training program. The sample consisted of 264 police officer applicants from 25 law enforcement agencies across four different states. The sample was divided equally into two groups: those who passed and those who failed the police academy training program. The failed group consisted of those individuals that did not pass, for one of several possible reasons, the eight to twelve month training program that consisted of two phases. Phase one consisted of academic training and phase two consisted of field training (e.g., on-the-job training activities). The majority of candidates failed the written test that was part of phase one. The remaining candidates withdrew after learning of the actual field training activities, were terminated because of poor performance, or were allowed to drop out in lieu of being terminated. The “completed training group” consisted of those candidates who successfully passed the academic and field training phases and graduated to become law enforcement officers.

The second study was designed to differentiate between officers who engaged in counterproductive work behavior resulting in formal disciplinary or departmental action and those officers who engaged in no such behavior (e.g., control group). The sample
consisted of 400 police officer applicants from 39 law enforcement agencies across four
different states. The sample was divided equally into two groups: those requiring
disciplinary or departmental action and those who were free from such actions by the
department. The “disciplinary group” consisted of officers who were formally
disciplined for counterproductive work behavior subsequent to hire. Counterproductive
work behaviors included excessive force, sexual misconduct, substance abuse,
insubordination, theft, lying, multiple motor vehicle violations, inappropriate verbal
conduct toward the public, multiple duty violations, and undue use of force. The “control
group” included officer who did not engage in the level of misbehavior above, but that
had one unfounded complaint and one justified use of force.

The predictors used in the study were the Wonderlic Personnel Test, which is a
measure of general mental ability; selected CPI scales (e.g., Responsibility, Socialization,
and Self-Control); life history information divided into three domains (e.g., work history,
drug history, and criminal history) by SMEs and collected through a personal history
questionnaire; structured interview; background investigation; and lastly, a veracity index
based on inconsistent responses to similar questions across different formats.

The authors conducted univariate analyses (F tests and effect sizes) assessing the
ability of the eight predictor variables to differentiate the dichotomous outcome variables.
In the first study, looking at candidates training failure or success, the results indicated
that the CPI personality variables failed to predict group membership. The work history
index of the life history indices was able to differentiate the two groups. The test of
mental ability was also predictive of candidate success in law enforcement training. In the
second study, concerning the prediction of officer misconduct on the job, seven of the

99
eight predictors were significantly correlated with the criterion. All three of the CPI personality variables were significantly related to group membership. Both the work and drug history indices were significantly related to the criterion.

The results from Cuttler and Muchinsky (2006) are similar to Sarchione et al. (1998) in terms of the predictive validity of the CPI variables that represent the conscientiousness construct when applied to counterproductive work behavior in law enforcement settings. However, the Cuttler et al. (2006) study differed from Sarchione et al. (1998) in that the drug history index was higher for the non-disciplined officer group. The second Cuttler et al. (2006) study also differed from Sarchione et al. (1998) in that two additional predictors (e.g., general mental ability and veracity index) were used and found significantly correlated with the criterion.

Limitations

There are several limitations that become apparent when examining a large body of published research on a particular topic. All studies have strength and weaknesses, and the previously reviewed work is no exception. Hiatt and Hargrave (1988) pointed out that most of the studies validating personality tests in the area of predicting law enforcement performance “have limited utility because of several methodological problems” (p. 122). Research in law enforcement, especially that research relating personality measures to job performance, is fraught with methodological limitations (Cuttler & Muchinsky, 2006; Hogan, Hogan, & Roberts, 1996).

Research examining the validity of personality measures in the prediction of police officer performance is confronted with the difficulty of obtaining large and meaningful police officer samples. Other concerns regarding the personality testing and
police performance research that reduce the robustness or generalizability of findings include “generating theoretically sound predictions, obtaining objective and less subjective measures of police performance, and utilizing comprehensive pre-hire measures of personality” (Enright, 2004, p. 20).

Ben-Porath (2003) noted that many of the studies are investigating the job performance of officers with an instrument that was used to screen the officers in the first place. Such validation studies do not allow for the assessment of performance of those applicants who would have been disqualified and therefore suffer from the statistical limitation of range restriction, which is considered an internal threat to validity (e.g., sampling bias).

An even more subtle manifestation of restriction of range in predictive validity studies has to do with the relationship of any mechanism, whether instrumental or procedural, to the instrument under investigation and the subsequent influence of that mechanism on the applicant pool. According to Hiatt and Hargrave (1988), other authors have gone as far as to propose that psychological evaluation may be unnecessary because other selection procedures such as the background investigation are successful at screening out those applicants who are psychologically unsuitable. However, Hiatt and Hargrave (1988) did not cite the source of this information. Further, Ben-Porath (2007) explained that most predictive validity studies in the area of employment screening are affected to some degree by range restriction. However, such threats to internal validity can be managed through the application of statistical correction methods. Despite the range restriction that occurs in law enforcement samples as a result of the relatively select group of individuals that complete the extensive screening and training processes, the
predictive validity of the CPI scales in the context of police performance is robust (Hargrave & Hiatt, 1989).

The selection of criterion variables represents another limitation in police research. Studies looking at law enforcement recruits and academy performance, which represent a significant share of the research on police selection, suffer from criteria-related problems. The specific problems in such studies is that the psychological variables required for successful completion of training may be different in many ways than those attributes necessary to be successful in the field (Hiatt & Hargrave1988). Problems related to outcome variables are considered an external threat to validity. The central issue is that the findings may not be relevant to actual on-the-job behavior and thus not generalizable.

Another limitation cited in the literature pertains to reliability. James et al, (1984) raised issue with the reliability of both supervisor and interview board ratings as criterion variables. The central issue cited by the authors involved the subjective and qualitative nature of such ratings and the subsequent lack of reliability not only between raters but across time as well. Wright et al (1990) reported that very few of the empirical studies looking at the relationship between psychological test scores and performance have used behaviorally anchored rating scales (BARS), which are considered the most acceptable instrument to assess performance.

The various methodological weaknesses addressed here are but a sample of the many ways in which the designs of the experiments limit the validity and reliability of the findings. There are several other ways in which the design of a study can jeopardize the value of the results. Researchers should stand on the shoulders of those scientists before
them and find new ways to address the problems that limit previous work. The architects of new experiments should then share their strategies and the improved results with the scientific community in the spirit of technological advancement. Such collaboration will manifest in better selection procedures, leading to higher quality law enforcement, and ultimately a more peaceful society.

Summary of the Results

Hargrave and Hiatt (1989) claimed that there are many studies examining the relationship between the CPI and various police related outcome measures. According to Groth-Marnat (2003), several scales on the CPI have been found related to police performance. As mentioned previously, Roberts and Johnson (2001) reported extensive empirical support for the use of the CPI scales in the selection of law enforcement officers.

In this section, the available empirical literature relevant to the use of the CPI in law enforcement settings was reviewed. It was noted that much more research exists, albeit unpublished, concerning the use of the CPI in evaluating the suitability and performance of police applicants and incumbents. Nevertheless, there is a substantial body of published work on the topic that provides support for the use of the CPI in identifying those police officer applicants who will likely manifest counterproductive work behavior (e.g., problems on the job). Based on the literature reviewed above, there are several CPI scales as well as supplemental scales or indices that have been repeatedly found both significantly related to and predictive of various criteria used to measure employment suitability of police officer applicants and job performance of incumbent law enforcement officers.
In terms of successful and unsuccessful applicants as well as both cadets and officers, the CPI scales of To, Wb, Ai, Ie, Re, Cm, So, Do, Sc, and Sp were consistently found to be significantly related to and, in many cases, predictive of employment suitability and future job performance (Cuttler & Muchinsky, 2006; Aamodt, 2004; Ones, Viswesvaran, Cullen, Drees, & Langkamp, 2003; Schneider, 2002; Sarchione, Cuttler, Mucinsky, & Nelson-Gray, 1998; Wells, 1991; Hargrave & Hiatt, 1989; Hiatt & Hargrave, 1988; Hargrave & Hiatt, 1987; Hartman, 1987; Hogan in Gough, 1987; Fitzgerald, 1986; Gettys & Elam, 1985; Hargrave, Hiatt, & Gaffney, 1986; Pugh, 1985; Hargrave & Berner, 1984; Lietner & Sedlacek, 1976; Hogan, 1971; Hogan, 1973; Hogan & Kutines, 1975; Mills & Bohannon, 1980; Hortsman, 1976; Sarchione, Cuttler, & Muchinsky, 1998). Although far less evidence exists for other scales, the CPI scales of Ac, Sy, Sa, Py, Fe, Fx, Em, Cs, have also been linked to police officer applicants’ and incumbents’ future success (e.g., employment suitability and job fitness) (Schneider, 2002; Wells, 1991; Hartman, 1987; Hogan in Gough, 1987; Hargrave et al., 1986; Fitzgerald, 1986; Hortsman, 1976; Gettys & Elam, 1985; Hargrave & Berner, 1984; Mills & Bohannon, 1980; Hogan & Kurtines, 1975). According to Hargrave and Hiatt (1989), there is no evidence for a relationship between scale Gi and performance.

Several authors have identified the Class II scales (e.g., Re, So, Sc, Gi, Cm, Wb, and To) as indicators of successful police performance (James, Campbell, & Lovegrove, 1984; Hargrave & Hiatt, 1989). Gough (1987) cited an unpublished study by Hogan showing that social skills as measured by factor 2 scales was most related to the successful performance of cadets. The empirical literature also indicates that the Class III Scales of Ai, Ac, and Ie are significantly related to and predictive of employment
suitability and job performance of law enforcement applicants and incumbent officers (Hargrave & Hiatt, 1989). Schneider (2002) also demonstrated that the Class III scale of Achievement via Conformance (Ac) was useful in employment selection. Conversely, the Class IV variables have not been found significantly related to policed suitability and fitness. Corey and Stewart (2007) found no significant relationship between Class I Scales and background investigator’s suitability determination. However, Schneider (2002) provided support for the use of several Class I scales (e.g., Responsibility, Socialization, Self-Control, Communality) in officer selection. Additionally, Gough (1987) cited an unpublished study by Hogan showing that Factor 1 scales that measure interpersonal values are most related to trooper performance.

A literature review conducted by Hargrave and Hiatt (1989) found no consistent results supporting a relationship between Class I or IV scales and police performance. However, the study by Corey and Stewart (2007) revealed significant findings in terms of the relationship between Class IV variables and background investigator’s suitability determination of police officer applicants. Lastly, the literature also reveals significant findings for certain supplemental scales. The CPI index scores of Lead, Wo, and PPE were found to be positive indicators of employment suitability and successful police performance (Hargrave and Hiatt, 1987; Hargrave, Hiatt, & Gaffney, 1986; Hargrave & Berner, 1984; Mills and Bohannon, 1980; Hogan & Kutines, 1975; Hogan, 1971).

According to Hargrave and Hiatt (1989) “One of the most import goals of screening applicants for law enforcement positions is the identification of characteristics that may contribute to major job problems” (p. 275). Most all of the CPI scales have been shown to be related to some degree to police officer job performance. Although some
scales are more related to police performance than others, certain scales, such as those of Gough’s Cluster II and III, have proven particularly useful in the forecasting of counterproductive work behavior. These scales have demonstrated predictive validity in the domain of psychological suitability and fitness of law enforcement applicants and incumbent police officers, respectively.

In addition to the scales that have already been established as useful in terms of predicting counterproductive work behavior, there are additional indicators that have been developed in the form of algorithms (e.g., risk estimates), which have not yet been systematically examined or well validated for the purposes of evaluating psychological suitability for police work. As indicated previously, Kostman (2004) looked at the correlation between CPI risk estimates from the “Job Suitability Snapshot” and police officer fitness for duty finding that the CPI risk ratings accurately predicted 67.4 percent of fitness categories…” (p. 89). The authors found that the CPI risk estimates “poorly suited” and Integrity were significant predictors. However, far more research is required in order to develop a solid scientific foundation upon which to validate the job suitability snapshot. The present research project was aimed to contribute to the research on the CPI risk estimates in law enforcement settings. The current study was designed to explore the predictive validity of the relatively newly developed CPI risk ratings in the context of police officer applicant’s suitability for law enforcement work.
THE CURRENT STUDY

The present study is an examination of the validity of the CPI when applied to the prediction of police applicant suitability as determined by background investigators. The particular design used a blind procedure in which no CPI data were used when forming judgments about applicant qualifications. Therefore the study design circumvents the single most pervasive limitation in personnel selection research-range restriction.

The study consists of 372 police officer applicants from a large northwest metropolitan police department within the United States. The sample was made up of two roughly equivalent and distinct groups of police officer applicants. The first group consisted of 201 applicants terminated for cause early in the screening process based on non-medical information (Integrity violations, recent illegal drug use, particular criminal offenses, poor work history, etc.). The second group consisted of 171 applicants who passed an exhaustive non-medical background investigation and received a conditional offer of employment. The current study examined the validity of the CPI in predicting membership in either the non-medically terminated (i.e., unqualified) group or the non-medically qualified group (COE).

In addition to gathering descriptive statistics and base rates for CPI suitability risk ratings, simple correlation analyses were conducted. Both single predictor and hierarchical logistic regression analyses were undertaken to determine the success of the CPI in predicting group membership in either the non-medically unqualified group or the group that was found qualified after a full-field background evaluation. Finally, overall correct classification was assessed as well as specificity and sensitivity. All analyses were conducted using Statistical Package for the Social Sciences (SPSS).
One issue that is particularly important regarding the proposed study is that the CPI is not a measure of psychopathology; but rather, a measure of normal personality (Gough, 1995). The non-medical nature of the CPI is important in that it is not prohibited by federal law (e.g., ADA) from the pre-offer phase of employment selection according to the Equal Employment Opportunity Commission. To the extent that the CPI is shown to have predictive validity in this context, the evidence may support the use of the CPI as a selection tool very early in the evaluation of police officer applicants, an outcome with potential widespread benefits.

Issues this Study is Tackling and Hypotheses

There are several issues that this study was designed to address. From a more broad and practical perspective, this study was intended to provide information that could increase the likelihood of selecting more suitable police officer applicants that would be successful in their job of serving the public and maintaining the peace. At a more specific level, the goal was to examine if using the CPI would result in a decreased rate of hiring errors, since such errors have manifold negative consequences with far reaching economic, human, and political implications. Additionally, the study aims to provide information regarding whether or not the CPI demonstrates economic utility if applied at the earliest hiring phase as a way of identifying those officers that would likely be terminated for cause later in the selection process.

From a scientific perspective, this study was intended to provide additional empirical data to the relatively small body of published work regarding the CPI and its relationship to the employment suitability of law enforcement applicants. The
dissemination of such information is essential for continued understanding and scientific progress in the area of psychological evaluation of police officer applicants. Considering the paucity of research in this field, even the most basic analysis of selected CPI scales and their relationship to preemployment screening criterion would prove to be a meaningful contribution to the literature (Corey, personal communication, 2007). Further, the near absence of published scientific work on the most recently developed feature of the CPI special police and public safety report, the eight Risk Estimates of the Job Suitability Snapshot, provided an opportunity to add to the empirical knowledge regarding these variables.

The present study was designed to examine the relationship between the Job Suitability Snapshot Risk Estimates of the CPI Police and Public Safety Selection Report and the dichotomous background investigation outcome of terminate for cause or COE. Additionally, the current research project was intended to examine the predictive validity of the eight risk estimate variables in terms of failing or passing the background investigation. The correlational aspect of the study will provide information about the nature of the variables from a statistical standpoint. Such information may facilitate a better understanding of what these variables are measuring and how clean of a construct each variable is. Information about the relatedness of the variables is particularly important in terms of the CPI because the issue of overlapping scales has been one of the most frequently and persistently cited critiques of the measure.

Lastly, information about the predictive validity of the CPI is essential for three reasons. First, there is a lack of information supporting the use of these variables in a preemployment selection context. Second, selecting suitable candidates, those who will
not engage in deviant work behavior (e.g., corruption, excessive use of force, discrimination, etc.), is one of the most critically important and complicated tasks undertaken by law enforcement agencies. Third, the non-medical nature of the CPI allows for its implementation very early in the selection process. Other personality tests are not allowed to be used prior to a conditional offer of employment because they are considered medical in nature and prohibited by federal regulations (See ADA, EEOC, and Leonel et al. v. American Airlines, Inc.). Therefore, information about the accuracy of the CPI in identifying which applicants are most likely to engage in antisocial (e.g., reckless and harmful) behavior, is extremely useful at the earliest stages where such information would not otherwise be accessible through other means. If the CPI is able to predict, with a reasonable degree of certainty, which applicants will be determined unsuitable for law enforcement work, than the measure can be used to weed out those individuals before a multitude of resources are wasted.

The current examination was undertaken with two central ideas in mind, which make up the hypotheses of the study. The first hypothesis was that the CPI risk ratings, which represent psychologist’s employment suitability determinations, the likelihood of an applicant being fired after hired, and several behavioral domains linked to employment problems in police work, are indeed related to suitability determinations made by background investigators, who are considered opinion experts by industrial and organizational psychology standards (Levy, 2006). The idea that law enforcement background investigators qualify as opinion experts is also supported in the police psychology literature (Baehr et al., 1968). The second hypothesis was that the CPI risk
ratings would successfully predict group membership in terms of which applicants passed or failed the background investigation.
METHOD

Subjects

The subjects for this study consisted of 372 police officer candidates applying at one large municipal police department in a major northwestern United States city between January 2001 and February 2007. All subjects were screened by the agencies background investigators according to formal personnel assessment and selection procedures. No prior psychological testing was reviewed.

The subject pool consisted of two distinct groups of police officer applicants. The first group consisted of 201 (54.0 percent) applicants terminated for cause early in the screening process (e.g., before the psychologist’s pre-offer suitability assessment). The second group consisted of 171 (46.0 percent) applicants that were screened prior to the implementation of the bifurcated protocol and therefore subjected to a full background investigation prior to being given a conditional offer of employment (i.e., an offer of employment contingent upon the results of medical examination, both psychological and physical).

All officers were required to be at least 21 years of age at the time of hire. However, officers were permitted to be under the age of 21 as long as they would be 21 years of age at the time of employment. The following demographic information was collected: age, sex, and race. Age of officers at the time of their application ranged from 20-55 years (Mean = 28.49 years, SD = 6.023). The subjects consisted of 308 men and 64 women. 80.6 percent of officers classified themselves as Caucasian/White with other subjects being distributed among Hispanic (3.8 percent), African-American/Black (5.6
percent), Asian or Pacific Islander (4.8), American Indian or Alaskan Native (3.2 percent), or Other racial/ethnic groups (1.9 percent).

Measures

The CPI data were generated post-selection decision and used as predictor variables. The CPI was not scored prior to the pre-offer suitability screening and was not used by the Personnel Division in determining applicant qualifications. The CPI data used as predictor variables were the risk estimates that are generated for the CPI (434) Police and Public Safety Report. Risk estimates are reported as a numeric probability generated from prediction equations based on logistic regression analysis.

There are eight categories representing the various kinds of job-related risk for which applicants receive risk estimates. The different categories for which individuals receive risk estimates are as follows: The first category is the applicant’s likelihood of being rated as “poorly suited” by psychologists with expertise in the practice of police and public safety selection. The suitability ratings are a result of several longitudinal studies conducted by Roberts and colleagues. Additionally, there are six “problem” behaviors considered job related (e.g., Job Performance, Integrity, Anger Management, Alcohol Use, Illegal Drug Use, and Substance abuse). The six composite variables that make up the problem behavior categories were derived from CPI scales and their relation to items on the Johnson, Roberts, & Associates Personal History Questionnaire (PHQ). Lastly, police officer applicants receive a risk estimate concerning the likelihood of the applicant’s eventual non-voluntary termination (being fired) if he or she is selected for employment. The last category, labeled Probability of involuntary departure, for which an
applicant receives a risk rating, was also developed from the longitudinal research referenced above.

The outcome variables for all recruits were whether they had been terminated for cause early in the selection process or given a conditional offer of employment after being found non-problematic for purposes of employment upon completion of a full background investigation. The criteria against which applicants were determined to be terminated for cause or advanced to the next phase are based on 10 critical job dimensions, which can be found in appendix A.

Procedure

All subjects in the study were administered the CPI prior to a hiring decision as one part of a multiple phase screening protocol. The CPI scantron forms were then sent to the consulting psychologist. After conducting the background investigation, investigators gave each applicant a rating that functioned as a recommendation either to advance the applicant or “not to proceed” with an applicant. The decision of whether to advance an applicant to the next phase or “not to proceed” was solely based on non-medical information. See appendix A for a list of the domains that were used in the determination of applicant’s suitability.

The non-processed CPI scantron forms for all subjects were obtained from the files of the police department’s consulting psychologist. The outcome data (e.g., whether applicants were terminated or given a condition offer of employment) was also obtained from the consulting psychologist’s database. The police chief was approached and asked to participate in the study. Participation consisted of providing data from personnel files
of subjects. The police chief chose to have the personnel data collected by employees of the department. All personnel information was kept strictly confidential and used only for the purposes of the study.

CPI scantron forms were sent to Johnson, Roberts and Associates, Inc. to be processed by computer scoring software. After the CPI scantron forms were processed and test scores were generated, the collected outcome data from the consulting psychologist’s archive and the personnel data from the agency personnel files were matched to personality test scores and labeled by number.

After gathering descriptive statistics, simple bivariate correlational analysis was conducted to examine the relationship between the risk estimates as well as the relationship of each risk estimate to the outcome variable of passing or failing the complete background investigation. Logistic regression analysis was used to examine the degree to which the continuous predictor variables (CPI risk estimates) successfully predicted the dichotomous criterion variable of non-select (those applicants terminated for cause early in the selection process) or pre-offer (applicants given a conditional offer of employment and advanced to the next phase of the screening process). The first step in the analysis involved entering each risk estimate individually to determine the predictive value of each variable separately. Next, logistic regression analysis was conducted on all eight of the risk estimates simultaneously in order to determine the added predictive value accounted for by each additional risk rating. Finally, overall correct classification was assessed as well as specificity and sensitivity. The analyses were conducted controlling for age, sex, and ethnicity. The data was reanalyzed without controlling for age, sex, and ethnicity.
RESULTS

Descriptive Statistics

Preliminary analyses of the data were undertaken to obtain descriptive statistics on the variables of interest. As indicated in the methods section, the outcome or criterion variable of proceed (pre-offer) or do not proceed (non-select) is categorical. The frequency data indicated that the pre-offer group consisted of 171 applicants, which was 46 percent of the total sample. There were 201 applicants in the non-select group constituting 54 percent of the total sample. For the purpose of this study, taking into account the relatively small sample size (although the sample size is sufficient in this study for the number of independent variables being used) and the intended statistical analyses (non-parametric or quasi-parametric statistical tests), the roughly equivalent group sizes are appropriate if not necessary.

The descriptive statistics including the mean, five percent trimmed mean, median, standard deviation, as well as the 25th and 75th percentile for each of the eight continuous predictor variables (CPI risk estimates) are presented in Table 1.
Table 1.
Means, Trimmed Means, Medians, Standard Deviations, and 25th and 75th Percentiles of CPI Risk Ratings

<table>
<thead>
<tr>
<th>Risk Ratings</th>
<th>Mean</th>
<th>Trimmed Mean</th>
<th>Median</th>
<th>SD</th>
<th>Percentile 25th</th>
<th>Percentile 75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly suited</td>
<td>28.25</td>
<td>27.10</td>
<td>24.00</td>
<td>15.21</td>
<td>16.25</td>
<td>37.00</td>
</tr>
<tr>
<td>Job performance</td>
<td>40.23</td>
<td>39.82</td>
<td>38.50</td>
<td>12.44</td>
<td>31.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Integrity</td>
<td>33.63</td>
<td>33.32</td>
<td>32.50</td>
<td>9.64</td>
<td>26.00</td>
<td>39.00</td>
</tr>
<tr>
<td>Anger management</td>
<td>42.35</td>
<td>41.99</td>
<td>41.00</td>
<td>13.59</td>
<td>32.00</td>
<td>52.00</td>
</tr>
<tr>
<td>Alcohol use concerns</td>
<td>19.82</td>
<td>19.38</td>
<td>18.00</td>
<td>8.55</td>
<td>13.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>14.64</td>
<td>13.98</td>
<td>13.00</td>
<td>8.05</td>
<td>9.00</td>
<td>19.00</td>
</tr>
<tr>
<td>Substance abuse proclivity</td>
<td>37.01</td>
<td>36.61</td>
<td>36.00</td>
<td>11.93</td>
<td>28.00</td>
<td>46.00</td>
</tr>
<tr>
<td>Probability of involuntary departure</td>
<td>11.70</td>
<td>11.09</td>
<td>10.00</td>
<td>6.19</td>
<td>8.00</td>
<td>14.00</td>
</tr>
</tbody>
</table>

The descriptive statistics in terms of the means, medians, and standard deviations indicate that the sample in the current study is not significantly different from the normative sample. Base rates for the prediction equations (e.g., risk rating), which have been calibrated and cross-validated on large samples of previous applicants (Roberts and Johnson, 2001), are as follows for Poorly suited, Job performance, Integrity, Anger management, Alcohol use concerns, Illegal drug use, Substance abuse proclivity, and Probability of involuntary departure: 25, 38, 27, 38, 16, 13, 33, 10, respectively. The mean scores for the individual risk estimates indicate that the sample population for this study scored higher on all of the risk ratings than the base rates for the normative sample. Although the five percent trimmed mean, which eliminates the extreme outliers, reflects lower average scores, albeit insignificantly lower, the scores for the sample in this study are still higher than the normative sample. Interestingly, the range of scores within each risk rating appeared to vary widely as indicated by the large standard deviations for each risk rating, which were in some cases more than half of the actual rating.
Based on the established risk level cut scores, the distribution of scores for each risk rating in the sample are all within the same risk level as those risk ratings for the normative sample. The risk ratings Alcohol use concerns, Illegal drug use, and Probability of involuntary departure, were all within the “Low” range in terms of level of risk, whereas the remaining risk estimates all fell within the “Moderate” level of risk range. The 25th and 75th percentiles for the risk estimates indicate that the distribution for each risk rating is not normal. The percentiles reflect a positive distribution.

Descriptive statistics were also explored in order to gather information concerning the distribution of scores on the continuous variables. The five percent trimmed mean does not appear to be “very different” from the original mean on any of the risk estimate scores, which means that some of the extreme scores are not having a strong influence on the original mean. The original mean and five percent trimmed mean for each risk rating is as follows: Poorly suited (28.25 and 27.10), Probability of involuntary departure (11.70 and 11.09), Substance abuse proclivity (37.01 and 36.61), Alcohol use concerns (19.82 and 19.38), Illegal drug use (14.64 and 13.98), Anger management (42.35 and 41.99), Integrity (33.63 and 33.32), and Job performance (40.23 and 39.82).

The Kolmogorov-Smirnov statistic provides information about the normality of the distribution of scores. Non-significant results, which are reflected in a $p$ value greater than .05, indicate normality. The results of the Kolmogorov-Smirnov statistic indicate that all $p$ values, except for that of the Anger management risk rating, were less than .05. Although such values suggest a violation of normality, this is often the case with larger samples. In this sample, the $p$ values were .000 for all risk ratings except Anger management.
Normality can also be assessed by obtaining skewness and kurtosis values. Both skewness and kurtosis values were examined for the variables of interest. Skewness values provide an indication of the symmetry of the distribution. Positive skewness values indicate positive skew (scores cluster to the left at the lower values). All of the predictor variables were positively skewed. Kurtosis values provide information about the peakedness of the distribution. Positive kurtosis values indicate that the distribution is rather peaked (clustered in the center), with long thin tails. There were positive kurtosis values for the following risk ratings: Poorly suited, Probability of involuntary departure, Alcohol use concerns, Illegal drug use, and Integrity. Kurtosis values below zero indicate a distribution that is relatively flat (too many cases in the extremes). There were three risk ratings (Substance abuse proclivity, Anger management, and Job performance) that had values below zero. However, according to Pallant (2001), with reasonably large samples of 200 cases or greater, as is the case in the present study, neither skewness nor kurtosis will have a significant impact on the results of analyses.

In order to further assess for normality, the shape of the distributions for each risk rating were examined using a histograms. Based on the shape of the histograms, the distributions of scores appear positively skewed. The Normal Q-Q Plots were examined for each risk rating as well. The observed value for most risk ratings appeared to reflect a departure from the expected value from the normal distribution. Also, the Detrended Normal Q-Q- Plots were examined and revealed a departure from the zero line indicating a deviation, as most data points should collect around the zero line.

Finally, boxplots for the distribution of scores for all risk ratings were obtained to assess for outliers. SPSS considers data points that extend 1.5 box-lengths from the edge
of the box to be outliers and those data points extending more than three box-lengths to be extreme outliers (Pallant, 2001, p. 61). An examination of the boxplots revealed that there were outliers on every risk rating. There were 12 outliers found for the ‘Probability of involuntary departure’ risk rating, three of which were extreme. There were nine outliers on the ‘Illegal drug use’ risk rating, none of which were extreme. There were seven outliers found for the ‘Poorly suited’ risk rating, one of which was extreme. There were four outliers found on the ‘Integrity’ risk rating and none of them were extreme. There were three outliers found on the Alcohol use concerns’ risk rating. The remaining risk ratings of Substance abuse proclivity, Anger management, and Job performance all contained one outlier. The outliers will be looked at later when prediction is discussed.

Taking all of the descriptive data into account, it is apparent that the distributions of scores for all of the risk ratings are positively skewed. Additionally, there are outliers on all risk ratings with two risk ratings (Probability of involuntary departure and Poorly suited) containing extreme outliers. In some studies where outliers appear to be impacting the distribution significantly, extreme scores may be transformed statistically or removed altogether. In the current study, an examination of the difference between the five percent trimmed mean and the original mean reveal that the two values are not significantly different, which indicates that the outliers will not significantly affect the statistical analyses. Therefore, the scores were retained in the data file.

Lastly, although the Kolmogorov-Smirnov test of normality reveal that most of the risk ratings violate the assumption of normality, this is common in larger samples of 200 or more cases. More importantly, the assumption of normality is only necessary when conducting parametric statistics. In non-parametric statistics, the most important
issue is having roughly equivalent groups, which is the case in the present study. Additionally, because the dependent/criterion variable is dichotomous, different distributional assumptions are made and normality is undesirable. Further, some applied statisticians point out that outliers are often the cases of most interest, which is particularly accurate in the present study (Thomas, personal communication, 2007).

Cases near the mean are not the ones of most interest when studying an instrument used to assess for individuals who may be likely to display deviant behavior. The extreme cases are of most interest because these cases are the ones that create problems for the hiring law enforcement agency.

**Correlation Analyses**

A simple bivariate correlational analysis was conducted to evaluate the relationship between all of the variables in the study. The correlations among the eight predictor variables are presented in Table 2. The results show that many of the predictor variables are highly correlated with each other. Tabachnick and Fidell (1996) suggest omitting or creating composite variables from those variables with a correlation coefficient of .7 or more. According to Grimm and Yarnold (1995) correlations of $r > .80$ between predictors should be considered very problematic. Pallant (2004) considers an $r$ of .9 and above to indicate that independent variables are highly correlated. However, Pallant (2004) also endorses Tabachnick and Fidell’s (1996) suggestion. Additionally, the ‘Collinearity Diagnostics’ generated from SPSS reveal that many of the ‘Tolerance’ values for the predictor variables were very low indicating that the multiple correlation with other variables is high, which suggests the possibility of multicollinearity. The
existence of multicollinearity is a violation of one assumption of logistic regression analysis. The only variable that does not appear to violate the assumption of multicollinearity is ‘Probability of involuntary departure’ with a Collinearity Tolerance statistic of .446. All other predictor variables have a Collinearity Tolerance statistic of less than .3 indicating high correlation with other variables other than the criterion. According to Grim and Yarnold (1995), “… the greater the multicollinearity, the more problems exist in terms of technical aspects (e.g., mathematical solutions and statistical inference), as well as for practical prediction and theoretical interpretations” (p. 45).

Table 2.
Correlations Among Predictor Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Poorly suited</td>
<td>-</td>
<td>.670</td>
<td>.853</td>
<td>.749</td>
<td>.763</td>
<td>.689</td>
<td>.512</td>
<td>.745</td>
</tr>
<tr>
<td>4. Illegal drug use</td>
<td></td>
<td></td>
<td>-</td>
<td>.750</td>
<td>.750</td>
<td>.601</td>
<td>.883</td>
<td></td>
</tr>
<tr>
<td>5. Anger management</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>.856</td>
<td>.826</td>
<td>.853</td>
<td></td>
</tr>
<tr>
<td>6. Integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>.770</td>
<td>.813</td>
<td></td>
</tr>
<tr>
<td>7. Alcohol use concerns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>.816</td>
<td></td>
</tr>
<tr>
<td>8. Substance abuse proclivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Logistic Regression Analyses

A series of logistic regression analyses were then conducted. First, logistic regression analyses were conducted on each of the CPI risk ratings. For each analysis, the predictor variable was entered into the equation in the second block in order to control for
ethnicity, sex, and age. The results of the logistic regression analyses for each risk rating entered separately after controlling for ethnicity, sex, and age, are presented in Table 3. A preliminary review of the results from the regression analyses reveal that the predictor variables ‘Poorly suited’ and ‘Probability of involuntary departure’ show a statistically significant and moderate relationship (-.397 and -.378, respectively) with the dichotomous (proceed/do not proceed) criterion variable of background investigation outcome. The predictor variables, ‘Job performance’, ‘Illegal drug use’, ‘Substance abuse proclivity’, ‘Anger management’, and ‘Integrity’ show a significant but weak correlation with the criterion variable (-.271, -.267, -.226, -.205 and -.188, respectively). Lastly, the ‘Alcohol use concerns’ variable was not statistically significant.

Logistic regression analysis also provides information indicating the odds ratio between the predictor variable and the criterion. The odds ratio “is the increase (or decrease if the ratio is less than one) in odds of being in one outcome category when the value of the predictor increases by one unit” (Tabachnick & Fidell, 1996, p. 607). When reviewing the results it is important to remember that the predictors represent estimated risk for the problem behavior that the particular risk rating was intended to measure. Therefore, a lower score would be hypothesized to represent a decreased risk in the behavioral domain (e.g., integrity) for which the predictor (e.g., Integrity) measures, and thus a lower likelihood that the applicant will engage in counter productive work behavior within that domain (e.g., taking a bribe). Accordingly, a higher risk estimate on a particular predictor (e.g. Illegal drug use) reflects an increased likelihood that the applicant will engage in problematic work behavior (e.g., the use or distribution for profit of controlled substances or the abuse of prescription medication such as using pain pills
for reasons other than they were prescribed) related to the domain (e.g., illicit substance use) for which that predictor is hypothesized to measures.

The results of the logistic regression analyses for each risk rating entered separately in terms of OR are as follows: For every one point change in the variable ‘Probability of involuntary departure’, the probability of an applicant being in the ‘do not proceed’ group increase by a multiplicative factor of 1.22. The inverse interpretation is that the probability of being in the ‘proceed’ group decrease by a factor of .804 for every one unit increase in the predictor ‘Probability of involuntary departure’. The odds of an applicant being in the ‘do not proceed’ group increase by a multiplicative factor of 1.08 and 1.07, respectively, for every one point change in the variables ‘Illegal drug use’ and ‘Poorly suited’. Again, the inverse interpretation is that for every one unit increase in the predictors ‘Illegal drug use’ and ‘Poorly suited’, the probability of being in the ‘proceed group’ decreases by a multiplicative factor of .924 and .932, respectively. The odds of an officer being in the ‘do not proceed’ group increase by a factor of 1.05, 1.05, and 1.04, respectively, for every one point change in the variables ‘Job performance’, ‘Integrity’, and ‘Substance abuse proclivity’. The probability of being in the ‘proceed’ group decreases by a factor of .952, .956, and .958, respectively, for every one unity increase in the predictors. For every one point change in the variable ‘Anger management’, the probability of being in the ‘proceed’ group decreases by a multiplicative factor of .967. That means that for every one unit increase in the predictor ‘Anger management’ the odds of being in the ‘do not proceed’ group increase by a multiplicative factor of 1.03. Lastly, the relationship between the ‘Alcohol use’ variable and the criterion was not statistically significantly. A change in the ‘Alcohol use concerns’ variable resulted in
only about a 1.00 factor increase in the odds of an officer being in the ‘do not proceed’
group. That is, the odds of being in the ‘proceed’ group decrease by a multiplicative
factor of .987. When the odds are one, there is no effect-no change in odds.

Table 3.
Logistic Regression Analysis of Background Investigation Outcome on CPI Variables
(After Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>CPI Risk Ratings</th>
<th>Correlation</th>
<th>B</th>
<th>SE</th>
<th>Wald(df=1)</th>
<th>P</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly suited</td>
<td>-.397</td>
<td>-.071</td>
<td>-.010</td>
<td>46.656</td>
<td>.000</td>
<td>.932</td>
<td>.913- .951</td>
</tr>
<tr>
<td>Job performance</td>
<td>-.271</td>
<td>-.049</td>
<td>-.010</td>
<td>25.012</td>
<td>.000</td>
<td>.952</td>
<td>.934- .971</td>
</tr>
<tr>
<td>Integrity</td>
<td>-.188</td>
<td>-.045</td>
<td>-.012</td>
<td>13.712</td>
<td>.000</td>
<td>.956</td>
<td>.933- .979</td>
</tr>
<tr>
<td>Anger management</td>
<td>-.205</td>
<td>-.034</td>
<td>-.009</td>
<td>15.389</td>
<td>.000</td>
<td>.967</td>
<td>.951- .983</td>
</tr>
<tr>
<td>Alcohol use concerns</td>
<td>-.036</td>
<td>-.013</td>
<td>-.013</td>
<td>1.066</td>
<td>.302</td>
<td>.987</td>
<td>.962- 1.012</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>-.267</td>
<td>-.079</td>
<td>-.016</td>
<td>23.437</td>
<td>.000</td>
<td>.924</td>
<td>.895- .954</td>
</tr>
<tr>
<td>Substance abuse proclivity</td>
<td>-.226</td>
<td>-.043</td>
<td>-.010</td>
<td>18.605</td>
<td>.000</td>
<td>.958</td>
<td>.939- .977</td>
</tr>
<tr>
<td>Probability of involuntary departure</td>
<td>-.378</td>
<td>-.218</td>
<td>-.032</td>
<td>45.577</td>
<td>.000</td>
<td>.804</td>
<td>.755- .857</td>
</tr>
</tbody>
</table>

Note: B = unstandardized coefficient, OR = odds ratio, CI = confidence interval.

Next, the logistic regression analyses previously conducted on the CPI risk ratings
were repeated. However, the second single predictor logistic regression analyses were
conducted without controlling for ethnicity, sex, and age. The results of the logistic
regression analyses for each CPI risk rating entered separately without controlling for
ethnicity, sex, and age, are presented in Table 4. A preliminary review of the results from
the regression analyses indicate no significant difference in the odds ratios from that
observed in the previous logistic regression analysis of the CPI risk ratings when controlling for ethnicity, sex, and age. All B values were negative indicating that as the CPI risk estimates increase, the odds of being in the ‘advance/proceed’ group (e.g., the predicted odds) decrease. The odds of an applicant being in the ‘do not proceed’ group increase by a multiplicative factor of 1.23, 1.08, 1.07, 1.05, 1.04, 1.04, 1.03, and 1.00 for every one point increase in the predictors Probability of involuntary departure, Illegal drug use, Poorly suited, Job performance, Integrity, Substance abuse proclivity, Anger management, and Alcohol use concerns, respectively. The results of the analyses indicate that the CPI risk rating do just as well at predicting group membership without controlling for ethnicity, sex, and age.

Table 4.
Logistic Regression Analysis of Background Investigation Outcome on CPI Variables (Without Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>CPI Risk Ratings</th>
<th>B</th>
<th>SE</th>
<th>Wald(df=1)</th>
<th>P</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly suited</td>
<td>-.070</td>
<td>.010</td>
<td>49.014</td>
<td>.000</td>
<td>.933</td>
<td>.915-.951</td>
</tr>
<tr>
<td>Job performance</td>
<td>-.048</td>
<td>.009</td>
<td>25.678</td>
<td>.000</td>
<td>.953</td>
<td>.936-.971</td>
</tr>
<tr>
<td>Integrity</td>
<td>-.041</td>
<td>.012</td>
<td>12.726</td>
<td>.000</td>
<td>.960</td>
<td>.938-.982</td>
</tr>
<tr>
<td>Anger management</td>
<td>-.032</td>
<td>.008</td>
<td>15.080</td>
<td>.000</td>
<td>.969</td>
<td>.953-.984</td>
</tr>
<tr>
<td>Alcohol use concerns</td>
<td>-.009</td>
<td>.012</td>
<td>.485</td>
<td>.486</td>
<td>.992</td>
<td>.968-1.016</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>-.076</td>
<td>.016</td>
<td>24.107</td>
<td>.000</td>
<td>.927</td>
<td>.899-.955</td>
</tr>
<tr>
<td>Substance abuse proclivity</td>
<td>-.040</td>
<td>.009</td>
<td>18.130</td>
<td>.000</td>
<td>.960</td>
<td>.943-.978</td>
</tr>
<tr>
<td>Probability of involuntary departure</td>
<td>-.225</td>
<td>.032</td>
<td>49.303</td>
<td>.000</td>
<td>.799</td>
<td>.750-.850</td>
</tr>
</tbody>
</table>

Note: B = unstandardized coefficient, OR = odds ratio, CI = confidence internal.
After each CPI risk rating was entered separately, all of the predictors were simultaneously entered into the logistic regression analysis. The variables were entered based on their interrelationship with other variables. The predictors with the least relationship to other variables were entered first, with those more highly correlated being entered thereafter. The ‘Alcohol use’ variable was omitted from the analysis because it failed to reach statistical significance. Additionally, The Substance abuse proclivity variable was omitted from the analysis because it is a composite of the Illegal Drug Use and Alcohol Use Concerns variables and is highly correlated with these other variables resulting in matrix singularity. The results of the logistic regression analyses for all remaining risk ratings entered simultaneously after controlling for ethnicity, sex, and age, are presented in Table 5.

A preliminary review of the results from the regression analyses revealed that the only predictor variables that reached statistical significance were the ‘Probability of involuntary departure’ and ‘Poorly suited’ variables. For every one point change in the variable ‘Probability of involuntary departure’, the odds of an applicant being in the ‘do not proceed’ group changed by a multiplicative factor of 1.13. For every one point change in the variable ‘Poorly suited’, the odds of an applicant being in the ‘do not proceed’ group change by a factor of 1.08. All other variables failed to reach statistical significance. One explanation for these results regarding the remaining variables, which are inconsistent with the results from the logistic regression analyses of individual risk ratings, is that the existence of multicollinearity was great enough to spoil the statistical analysis resulting in inaccurate results for those variables.
The results indicate that both the ‘Probability of involuntary departure’ and ‘Poorly suited’ variables are significant predictors and that change in these variables reflect a change in the odds that an individual will belong to the ‘do not proceed’ group. The results regarding the ‘Probability of involuntary departure’ and ‘Poorly suited’ variables from the hierarchical logistic regression analysis are consistent with the findings from the logistic regression analysis looking at individual risk ratings. As indicated previously, the only variable that does not appear to violate the assumption of multicollinearity is ‘Probability of involuntary departure’. Additionally, the ‘Probability of involuntary departure’ variable was correlated with the ‘Poorly suited’ variable, although they did not meet the threshold of “high” according to any of the previously cited authors. Further, the results of the hierarchical regression analysis yielded B values for the two variables that, when combined, equal the B value for the ‘Probability of involuntary departure’ when entered individually in the first regression analysis conducted. One explanation for these results is that the variable ‘Probability of involuntary departure’ simply measures that which is most predictive in the ‘Poorly suited’ variable. If that is the case, than the ‘Probability of involuntary departure’ risk estimate is the most valuable predictor of all the risk ratings. That is, the odds of being in the ‘do not proceed’ group increase by a factor of 1.21 with every incremental increase in the predictor ‘Probability of involuntary departure’.
Table 5. Hierarchical Logistic Regression Analysis of Background Investigation Outcome on CPI Variables (After Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>CPI Risk Ratings</th>
<th>B</th>
<th>SE</th>
<th>Wald(df=1)</th>
<th>P</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of involuntary</td>
<td>-.137</td>
<td>.039</td>
<td>12.182</td>
<td>.000</td>
<td>.872</td>
<td>.807-.942</td>
</tr>
<tr>
<td>departure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrity</td>
<td>.000</td>
<td>.028</td>
<td>.000</td>
<td>.995</td>
<td>1.000</td>
<td>.946-1.057</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>-.061</td>
<td>.040</td>
<td>2.367</td>
<td>.124</td>
<td>.941</td>
<td>.870-1.017</td>
</tr>
<tr>
<td>Poorly suited</td>
<td>-.088</td>
<td>.026</td>
<td>11.691</td>
<td>.001</td>
<td>.916</td>
<td>.871-1.063</td>
</tr>
<tr>
<td>Job performance</td>
<td>.048</td>
<td>.032</td>
<td>2.219</td>
<td>.136</td>
<td>1.049</td>
<td>.985-1.117</td>
</tr>
<tr>
<td>Anger management</td>
<td>.037</td>
<td>.023</td>
<td>2.562</td>
<td>.109</td>
<td>1.038</td>
<td>.992-1.087</td>
</tr>
</tbody>
</table>

Note: B = unstandardized coefficient, OR = odds ratio, CI = confidence internal

The odds ratio (OR) “estimates the change in the odds of membership in the target group for a one-unit increase in the predictor.” (Grimm & Yarnold, 1995, p. 223). In tables 3 and 5, looking at logistic regression analyses of CPI risk ratings separately and looking at all risk rating simultaneously, B (e.g., the unstandardized coefficient) represents b₁ (e.g., the raw coefficient of the predictor variable). Although B (b₁) is more difficult to interpret than an odds ratio, the raw coefficient has a useful function. A positive predictor coefficient means that the predictive odds increase as the predictor values increase whereas a negative coefficient indicates that the predicted odds decrease as the predictor increases. In the analyses where all risk ratings were analyzed separately, all B values were negative indicating that as the CPI risk estimates increase, the odds of being in the ‘advance/proceed’ group (e.g., the predicted odds) decrease.
Additionally, according to Grimm and Yarnold (1995), a coefficient of zero means that there is no effect (e.g., no change in odds) and that the predicted odds are the same for any value of the predictor. In the analysis where all risk ratings were entered separately, the B values were less than one and closer to zero, which indicates that the OR is near one and the predicted odds are similar for any value of the predictor. For the logistic regression analyses looking at all risk ratings independently, the OR for each risk rating was about .9 indicating little change in odds. Although the negative raw coefficients for the predictor variables indicate that a one unit increase in risk ratings reflect a decreased probability of being in the “advanced/proceed” group (e.g., an increased probability of being in the ‘do not proceed’ group), the near zero B/ b₁ values also reflect that the change in odds are negligible.

In the hierarchical logistic regression analysis, when all variables were entered simultaneously, all B values for the CPI risk ratings were near zero indicating that, for those risk ratings, the predicted odds are similar for any value of the predictor. The OR values for ‘Integrity’, ‘Job performance’, and ‘Anger management’ all equaled one, indicating that both outcomes are equally likely. The other risk ratings have OR values of less than one, albeit only slightly, indicating that the target event (e.g., advance) is less likely than the other event (do not proceed). That is, the predicted odds decrease with an increase in the predictor. ORs “indicate how much more likely it is that an observation is a member of the target group rather than a member of the other group.” (Grimm & Yarnold, 1995, p. 223). Based on the results from the hierarchical regression analysis, one would conclude that the risk ratings are not very helpful in predicting which group an
applicant would belong to. The OR values are very similar for the first analysis, when the risk ratings were entered independently.

The value reported for the maximum likelihood parameter estimates in this study was -2 Log Likelihood (-2LL) or the deviance, which is the measure computed from the sample likelihood. The -2LL values for the single predictor models ranged from about 430 to about 490. For the model when all risk ratings were entered simultaneously, the -2LL value was 402.895. Smaller values on the -2LL indicate that the data fits the model better, whereas larger values indicate a poor-fitting model (Mertler & Vannatta, 2005). A perfect model has a value equal to zero for this measure of model fit (George & Mallery, 2000). Therefore, the values reported on this measure of model fit for the present study indicate poor-fitting models.

As a part of the regression analysis, hypotheses tests were performed to assess how well the models fit (e.g., goodness-of-fit). The likelihood ratio statistic is used to determine whether the prediction coefficient is zero or if it differs from zero. In the case of models that contain a single predictor, according to Grimm and Yarnold (1995) “the probability for the likelihood ratio statistic is obtained from a chi-square distribution with 1 degree of freedom.” (p. 227). A large likelihood ratio statistic means that the population coefficient probably differs from zero. All single predictor models (e.g., logistic regression analysis with risk ratings entered individually) revealed small Chi-square values around eight and failed to reach statistical significance at the .05 level. The results indicate that of the eight different prediction models, each consisting of individual risk ratings, none of the models fit the data well. Thus, the individual risk ratings do not appear to differentiate which group an applicant will belong to at a level of statistic
significance. In terms of the logistic regression model analyzing all six risk ratings simultaneously, the likelihood ratio statistic or Chi-square was 12.112 with 8 df and a $p$ value of .146. Unlike linear regression analysis, a high $p$ value is desirable, indicating that the data adequately fit the model and that no further parameters need to be estimated. Although the $p$ value was above .05, it remains relatively low indicating that it is unlikely that the data fit the model.

In addition to interpreting the likelihood ratio statistics to assess the ‘goodness-of-fit’ for the different models, confidence intervals were evaluated as well. Confidence intervals can also be used for hypothesis testing. The OR confidence interval for the current model was set at 95%. In the hierarchical logistic regression analysis, for all variables except ‘Probability of involuntary departure’ and ‘poorly suited’, the 95% confidence intervals contain the number one, indicating that the odds ratio is nonsignificant at the .05 level of significance. The results suggest that a change from one unit to another does not reliably increase the odds of membership in the target group. In the single predictor models where each predictor was entered into the logistic regression equation individually, the only risk rating that contained a one in the OR confidence interval was ‘Alcohol use concerns’, indicating a nonsignificant OR at the .05 level of significance.

Classification Analyses

Another method of assessing the success of a model according to Tabichnick and Fidell (1996) “…is to evaluate its ability to predict correctly the outcome category for cases for whom outcome is known” (p. 606). Classification tables provide a summary of
the fit between the actual and predicted group membership. The classification table for the hierarchical logistic regression model is shown in Table 6. In the classification table, the number of cases in the left to right diagonal cells (the failed-failed and passed-passed cells) is large compared with the number of cases in the other cells (off-diagonal). The pattern indicates a good match between the observed outcomes and those predicted by the model. The overall percentage of cases correctly classified by the model, that is, the percentage accuracy in classification, or PAC, is 74.2 percent.

Four other measures of classification accuracy were computed. The sensitivity, which is the percentage of the target (proceed) group accurately classified, was 70 percent; this is the correct identification of true positives. Conversely, 30 percent of those who were predicted to pass actually failed (e.g., false positives). The positive predictive value, that is, the percentage of individuals that the model classifies as belonging to the target group that are actually in the target group, was 78 percent. The specificity, which is the percentage of the other group that is correctly classified, also known as the “correct identification of true negatives”, is 79 percent. Conversely, 21 percent of those who were predicted to fail actually passed (e.g., false negatives). Lastly, negative predictive value is the percentage of individuals that the model classifies in the other group that are actually in the other group. The negative predictive value was 71 percent.

It is clear from the classification table and the calculation of both the sensitivity and specificity of the model, that the CPI risk ratings are more accurate in predicting group membership of those applicants who failed the background investigation than those who passed. Based on the classification analysis, 143 of the 372 applicants were identified by the CPI risk ratings as failed who actually failed the background check. In
order to eliminate from the applicant pool those applicants who will fail the background investigation, 38 applicants who would otherwise pass the background check would be eliminated from consideration. The implications of the classification analysis will be addressed further in the discussion section.

Table 6.
Classification Table for Hierarchical Logistic Regression Analysis of Background Investigation on CPI variables (After Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>Observed Background Investigation Outcome</th>
<th>Predicted Background Investigation Outcome</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed preliminary background investigation</td>
<td>Failed</td>
<td>143</td>
</tr>
<tr>
<td>Passed full background investigation</td>
<td>Passed</td>
<td>133</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The cut value is .500

The prediction model was reevaluated without controlling for ethnicity, age, and sex. The classification table for the hierarchical logistic regression model without controlling for ethnicity, age, and sex is shown in Table 7. As was the case with the classification analysis controlling for age, sex, and ethnicity, the pattern indicates a good match between the observed outcomes and those predicted by the model. However, the overall percentage of cases correctly classified by the model was 72.3 percent, which is slightly lower than the classification accuracy for the model when controlling for the demographic variables.
The *sensitivity*, which is the percentage of the target (proceed) group accurately classified, was 67 percent; this is the correct identification of true positives. Conversely, 33 percent of those who were predicted to pass actually failed (e.g., false positives). The *positive predictive value*, that is, the percentage of individuals that the model classifies as belonging to the target group that are actually in the target group, was 77 percent. The *specificity*, which is the percentage of the other group that is correctly classified, also known as the “correct identification of true negatives”, is 78 percent. Conversely, 22 percent of those who were predicted to fail actually passed (false negatives). Lastly, *negative predictive value* is the percentage of individuals that the model classifies in the other group that are actually in the other group. The negative predictive value was 68 percent.

Again, it is clear from the classification table and the calculation of both the sensitivity and specificity of the model, that the CPI risk ratings without controlling for demographic variables are more accurate in predicting group membership of those applicants who failed the background investigation than those who passed. The classification table predicting membership for the hierarchical regression analysis without controlling for ethnicity, age, and sex reveals that 173 individuals of the 372 applicants were correctly identified by the test who actually would fail the background check. To get them out of the system, 39 individuals who would have passed the background check would be eliminated from consideration. The implications of the classification analysis will be addressed further in the discussion section.
Table 7.
Classification Table for Hierarchical Logistic Regression Analysis of Background Investigation on CPI variables (Without Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>Observed Background Investigation Outcome</th>
<th>Predicted Background Investigation Outcome</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed preliminary background investigation</td>
<td>Failed</td>
<td>173</td>
</tr>
<tr>
<td>Passed full background investigation</td>
<td>Passed</td>
<td>39</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The cut value is .500

Classification information for the single predictor regression models when controlling for ethnicity, age, and sex is shown in table 8. Table 8. provides information regarding the PAC, sensitivity, specificity, positive predictive value, and negative predictive value for the individual risk estimates. The PAC for the single predictor models controlling for ethnicity, age, and sex was 71.2 percent for ‘Probability of involuntary departure’, 67.7 percent for ‘Poorly suited’, 60.8 percent for ‘Job performance’, 61.8 percent for ‘Illegal drug use’, 64.2 percent for ‘Substance abuse proclivity’, 61.6 percent for ‘Anger management’, 59.7 percent for ‘Integrity’, and 54.3 percent for ‘Alcohol use concerns’. Based on the sensitivity and specificity values as well as the positive and negative predictive values for the classification tables, it appears that the CPI risk estimates are slightly better at predicting those individuals that will fail the background investigation and thus is better at screening out unsuitable police officer
applicants than identifying those applicants who will be successful in the background screening and given a conditional offer of employment.

Table 8.
Classification Values for the Single Predictor Logistic Regression Analyses of Background Investigation on CPI variables (After Controlling for Ethnicity, Age, and Sex).

<table>
<thead>
<tr>
<th>Risk estimates</th>
<th>PAC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of involuntary departure</td>
<td>71.2</td>
<td>67.7</td>
<td>74.4</td>
<td>71.3</td>
<td>71.1</td>
</tr>
<tr>
<td>Poorly suited</td>
<td>67.7</td>
<td>64.2</td>
<td>70.9</td>
<td>67.3</td>
<td>68.2</td>
</tr>
<tr>
<td>Job performance</td>
<td>60.8</td>
<td>57.6</td>
<td>63.1</td>
<td>55.0</td>
<td>65.7</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>61.8</td>
<td>58.1</td>
<td>65.2</td>
<td>60.8</td>
<td>62.7</td>
</tr>
<tr>
<td>Substance abuse proclivity</td>
<td>64.2</td>
<td>61.8</td>
<td>66.0</td>
<td>57.9</td>
<td>69.7</td>
</tr>
<tr>
<td>Anger management</td>
<td>61.6</td>
<td>58.8</td>
<td>63.5</td>
<td>54.4</td>
<td>67.7</td>
</tr>
<tr>
<td>Integrity</td>
<td>59.7</td>
<td>56.7</td>
<td>61.7</td>
<td>51.5</td>
<td>66.7</td>
</tr>
<tr>
<td>Alcohol use concerns</td>
<td>54.3</td>
<td>50.2</td>
<td>58.8</td>
<td>57.9</td>
<td>51.2</td>
</tr>
</tbody>
</table>

The classification information for the single predictor regression models without controlling for ethnicity, age, and sex is shown in table 9. Table 9. provides information regarding the PAC, sensitivity, specificity, positive predictive value, and negative predictive value for the individual risk estimates. The PAC for the single predictor models without controlling for ethnicity, age, and sex was 68.5 percent for ‘Probability of involuntary departure’, 66.7 percent for ‘Poorly suited’, 61.3 percent for ‘Job performance’, 58.6 percent for ‘Illegal drug use’, 59.1 percent for ‘Substance abuse
proclivity’, 58.1 percent for ‘Anger management’, 58.3 percent for ‘Integrity’, and 54.0 percent for ‘Alcohol use concerns’.

Based on the sensitivity and specificity values as well as the positive and negative predictive values for the classification tables, it appears that the prediction models for the individual CPI risk estimates without controlling for ethnicity, age, and sex are not as good at predicting outcome (e.g., group membership) as the prediction models for the CPI risk estimates when those demographic variables were controlled for. Nevertheless, similar to the single predictor prediction models controlling for demographic variables, the prediction models without controlling for the demographic variables are better at predicting those who will fail the background check than those who will pass. Interestingly, the CPI risk rating ‘Alcohol use concerns’ demonstrated no improvement whatsoever over the prediction from base rates.
Table 9.
Classification Values for the Single Predictor Logistic Regression Analyses of
Background Investigation on CPI variables (Without Controlling for Ethnicity, Age, and
Sex).

<table>
<thead>
<tr>
<th>Risk estimates</th>
<th>PAC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of involuntary departure</td>
<td>68.5</td>
<td>64.2</td>
<td>73.0</td>
<td>71.3</td>
<td>66.2</td>
</tr>
<tr>
<td>Poorly suited</td>
<td>66.7</td>
<td>62.5</td>
<td>70.8</td>
<td>68.4</td>
<td>65.2</td>
</tr>
<tr>
<td>Job performance</td>
<td>61.3</td>
<td>58.0</td>
<td>63.9</td>
<td>56.7</td>
<td>65.2</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>58.6</td>
<td>54.7</td>
<td>62.1</td>
<td>57.3</td>
<td>59.7</td>
</tr>
<tr>
<td>Substance abuse proclivity</td>
<td>59.1</td>
<td>56.0</td>
<td>61.3</td>
<td>51.5</td>
<td>65.7</td>
</tr>
<tr>
<td>Anger management</td>
<td>58.1</td>
<td>55.6</td>
<td>59.4</td>
<td>43.3</td>
<td>70.6</td>
</tr>
<tr>
<td>Integrity</td>
<td>58.3</td>
<td>55.9</td>
<td>59.6</td>
<td>43.9</td>
<td>70.6</td>
</tr>
<tr>
<td>Alcohol use concerns</td>
<td>54.0</td>
<td>54.0</td>
<td>-</td>
<td>.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The final issue to be addressed in the current study involves those cases with extreme scores (e.g., outliers) on the predictors (e.g., risk estimates). As noted previously, outliers may be of particular interest in a study focused on screening for deviant behavior. Therefore it is important to look at what happens with the extreme scores in terms of prediction. Recall in the Results section under Descriptive Statistics that the five percent trimmed mean does not appear to be “very different” from the original mean on any of the risk estimate scores. Similar scores on the mean and five percent trimmed mean indicate that the extreme scores are not having a strong influence on the original mean and thus will not significantly influence the results of statistical analysis.
Another analysis was conducted to examine the possible effects of outliers. In order to understand whether or not the outliers had an effect on prediction, the primary hierarchical logistic regression analysis was conducted once more after the cases with extreme scores were omitted from the data set. The result of the hierarchical logistic regression analysis after omitting the outliers revealed no change in prediction. The classification table showed a pattern indicating a good match between the observed outcomes and those predicted by the model. The overall percentage of cases correctly classified by the model, that is, the percentage accuracy in classification, or PAC, is 74.0 percent. Based on the results of the prediction model after omitting the outliers, there is no reason to assume that the outliers in this study significantly affected the results of the statistical analyses conducted herein.

Lastly, the classification output for SPSS provides a list of outliers that are misclassifications. The extreme scores were cross-referenced with the list of misclassifications for the hierarchical logistic regression analysis and sigle predictor regression analyses. None of the cases that were outliers were among the list of misclassifications. Therefore, it does not appear that the prediction model, looking at all of the risk estimates simultaneously or individually, had any problem accurately predicting which group the cases with extreme scores actually belonged to. Such a finding is important because, as mentioned previously, cases near the mean are not the ones of most interest when studying an instrument used to assess for individuals who may be likely to display deviant behavior. The extreme cases are of most interest because these cases are the ones that create problems for the hiring law enforcement agency. The fact that the CPI risk estimates were successful in predicting group membership of those
individuals reporting an abnormally high number of behaviors considered problematic in employment settings, particularly safety sensitive positions, is support for the use of the CPI risk estimates in screening out potentially problematic applicants for the position of law enforcement officer.
DISCUSSION

Summary of Results

In the current study, it appears that police officer applicants as a group, whether they were advanced or not in the selection process after extensive background investigation, tend to endorse fewer items reflecting those problems that contribute to elevated risk ratings. It is common in the social sciences to have scales and measures that are skewed, either positively or negatively. These trends simply reflect the underlying nature of the constructs being measured rather than some problem with the measure itself (Pallant, 2004). The positive skew seen in this sample is consistent with preemployment selection samples in general (Corey, personal communication, 2007). According to Ben-Porath (2007), there is a degree of positive impression management that occurs in preemployment psychological evaluations. Applicants tend to put their best foot forward and present themselves in an unrealistically positive light, denying common human frailties. The phenomenon of positive impression management or “faking good” permeates the behavioral science literature related to psychological assessment in preemployment settings.

Although it is common for job applicants to attempt to appear especially “good” for purposes of securing employment, which is often a life time goal for many individuals seeking law enforcement positions, the issue of honesty is especially relevant for police officer applicants. Integrity is one of the most frequently referenced attributes in the police selection literature. High risk occupations such as law enforcement require honorable, organized, reliable, and well-regulated individuals to successfully discharge
the office of law enforcement officer. Therefore, the self-report of applicants requires extensive verification. Police officer applicants are well informed that their claims will be scrutinized and that deception as well as exaggeration is grounds for termination. The fact that police officer applicants are aware that their self-report will be checked for accuracy in many ways (e.g., reference checks, questionnaires to associates, and polygraph) likely results in increased honesty. Such accuracy in self-report would indicate that police officer applicants as a group simply engage in less deviant behavior. The most likely hypothesis is that the extensive employment selection procedures required for employment in law enforcement discourages those individuals who have a record of deviant behavior.

According to Roberts and Johnson (2001), each CPI risk rating is reported as a numerical probability of the undesirable outcome, which is categorized into three risk level categories: High risk (p > 50%), Moderate risk (p = 25%-49%) or Low risk (p < 24%)” (p. 15). The CPI special report also provides base rate values for the large number of police and public safety applicants that make up the normative sample. Based on the established risk level categories, the distribution of scores for each risk rating in the sample are all within the same risk level categories as those risk ratings for the normative sample. The risk ratings Alcohol use concerns, Illegal drug use, and Probability of involuntary departure, were all within the “Low” range in terms of level of risk, whereas the remaining risk estimates all fell within the “Moderate” level of risk range. Therefore, it appears that the current sample is similar to the normative sample in terms of risk ratings.
As noted above, there is a significant response bias evoked in applicants undergoing preemployment testing, which may be the cause of the positively skewed scores in the current study sample. An alternative explanation for the positively skewed scores is that the extensive and rigorous employment selection process for law enforcement officers may discourage those individuals with a history of documented deviant behavior, resulting in selection bias and subsequently creating an applicant pool that evidences very little problem behavior. Although there is extensive support in the selection literature for the former hypothesis, the current study may provide support for the latter.

Although six of the risk ratings come CPI scales related self-report of problem behavior, two of the risk ratings are based on the opinions of expert psychologists in the field of police psychology: ‘Poorly suited’ and ‘Probability of involuntary departure’. The sample distribution of scores for the two risk estimates that come from psychologist ratings were also positively skewed. The poorly suited risk estimate was developed using a sub-sample of 22,867 cases and involved a completed psychological evaluation of each applicant that included a face-to-face structured interview, the CPI, and other tests such as the PAI, MMPI, STAXI, as well as a personal history questionnaire, all resulting in a suitability rating: “suitable” and “poorly suited.” The Probability of involuntary departure risk estimate was the result of several longitudinal studies examining the predictive accuracy of the suitability determinations just mentioned. The studies yielded data indicating that certain classes of applicants within the “poorly suited” classification were three times more likely to be terminated for cause and two times more likely to engage in significant counterproductive behavior indicated by disciplinary reports.
The two risk estimates that represent determinations indicating either poor suitability or the probability of an applicant being fired if hired, from comprehensive psychological evaluation conducted by specially trained evaluators and subsequent validation studies confirming the accuracy of the evaluator’s predictions provide support for the hypothesis that police officer applicants as a group report very few norm violating behaviors. The chance that officer applicants are telling the truth rather than presenting themselves in an unrealistically positive light has implications for psychologists conducting preemployment psychological evaluations for police officer applicants. The implications are that this unique group of individuals perusing a career in law enforcement, may be telling the truth, when the results from personality testing indicate that they are being less than forthright in their self-report. It seems reasonable to conclude that the reality of the matter lies somewhere in the balance. It is most likely that the majority of police officer applicants are reasonably well socialized and principled individuals presenting themselves as “best” they can because they are highly invested in obtaining a position where they can serve the community and contribute to the peace.

In addition to the descriptive statistics, which provide information about what the sample of police officer applicants in this study look like, several additional analyses were conducted. Correlational analyses were undertaken to investigate the relationships of the variables to the criterion as well as the relationships among the variables. The results of simple correlation analyses conducted in this study indicate that both the ‘Probability of involuntary departure’ and ‘Poorly suited’ predictor variables demonstrated a statistically significant and moderate relationship with the background investigation outcome. The remaining predictors except for ‘Alcohol use concerns’ were
statistically significant, but only weakly correlated with the dichotomous criterion variable of passing or failing the comprehensive background investigation.

The ‘Alcohol use concerns’ risk estimate failed to demonstrate a significant relationship with the background investigation outcome. Initially, the nonsignificant finding regarding the ‘Alcohol use concerns’ predictor was thought to be attributed to the prohibition by the ADA to make inquiries regarding alcohol or drug use since such inquiries could reveal a disability. However, the ‘Illegal drug use’ predictor was found to be significantly related to the background investigation outcome, which violated the previous assumption. Further, investigation of the data set revealed that approximately 25 percent of the terminated group was eliminated for reasons related to drugs and alcohol. Unfortunately, the data was not coded in a manner that differentiated weather the reason was related to drugs or alcohol.

Nevertheless, background investigators are lawfully able to disqualify applicants for behavioral problems, even if those problems result from alcohol or drug abuse. For example, legal violations such as driving while under the influence of an intoxicant, possession of a controlled substance, and public intoxication, to name a few, are legal violations that may be used in determining an applicant’s suitability. Such legal violations are included as criteria in the investigator manual used by the agency in this study for making suitability determinations (See CREDIBILITY under Job Dimension # 10 STANDARDS OF CONDUCT as well as Driving Record under Job Dimensions # 2, 4, 6, and 8 in Appendix A.).

One explanation for the nonsignificant findings regarding the ‘Alcohol use concerns’ risk estimate has to do with the possibility that all of the applicants identified
as terminated for drug and alcohol related reasons were actually terminated only for problem behavior related to illegal substances. This hypothesis is based on the assumption that an insignificant number of applicants were terminated for reasons related to alcohol because such offenses were not considered egregious enough to warrant termination. One of the disqualifying criteria for police officer applicants at the hiring agency in this study is conviction of a crime, for which punishment would have been a felony in any jurisdiction. A felony is considered a serious crime, whereas misdemeanors are considered to be less serious offenses. Save domestic violence, most alcohol related offenses such as driving under the influence are treated as misdemeanor crimes in all states in the U.S. Only after a person’s third offense for driving under the influence is the alcohol related crime treated as a felony in the majority of U.S. states. Therefore, it is not likely that an alcohol related crime, or even several for that matter, save driving under the influence, would result in disqualification. Even in the case of driving under the influence, an individual must be convicted two or more times to be considered ineligible for employment at the represented agency.

Conversely, in terms of illegal drug use, the agency in this study considered conviction of any crime involving controlled substances a disqualifying behavior. Legally, an agency is able to disqualify applicants for admitted illegal drug use within a certain period of time. Background investigators are simply prohibited from inquiring about quantity or frequency of use since it could reveal a disability under the ADA. The zero tolerance for drug use at the represented agency is the best explanation for why most of the drug and alcohol related terminations were probably related to illegal substance use rather than alcohol related problems. Additionally, only one of the four items on the
‘Alcohol use problems’ risk rating is related to a legal violation (e.g., driving under the influence), whereas most all of the seven items on the ‘Illegal drug use’ risk rating involve illegal behavior not related to volume or frequency of use. The fact that Illegal substance use is an automatic disqualifying criterion, that most of those applicants identified by background investigators as being terminated for reasons related to drugs and alcohol were likely terminated for illegal substance use, and that most of the items on the CPI ‘Illegal drug use’ risk rating reflect illegal substance use seem to be reasonable explanations for the results indicating a much higher correlation between suitability determinations and the ‘Illegal drug use’ risk rating than the ‘Alcohol use concerns’ risk rating, which was nonsignificant.

The findings that all but the ‘Alcohol use concerns’ variables have a significant relationship with the outcome variable provides support to the already existing research that certain behavioral domains such as integrity, illegal drug use, emotional regulation (e.g., anger management), and past employment history are important in the determination of police officer suitability (Cuttler & Muchinsky, 2006; Black, 2000; Sarchione, Cuttler, Muchinskky, & Nelson-Grey, 1998). Additionally, the fact that there was a significant relationship between the variables ‘Probability of involuntary departure’ and ‘Poorly suited’, two variables derived from psychologist’s suitability ratings who were trained in police screening, indicate an agreement between psychologists suitability ratings and the ratings of trained background investigators. However, further studies designed specifically to examine the interrater reliability of psychologist’s and background investigator’s suitability determinations are necessary to validate these initial findings.
The simple correlational analyses looking at the relationship among the predictor variables yielded interesting results. Most of the correlations among the variables were high. The majority of the correlations between risk estimates were equal to or above .7. The correlations among risk estimates ranged from .150 between the variables ‘Probability of involuntary departure’ and ‘Alcohol use concerns’ to .883 between the Variables ‘Illegal drug use’ and ‘Substance abuse proclivity’. The risk rating ‘Probability of involuntary departure’ was the least related to any other risk estimates, with correlations ranging from .150 to .469 with the other variables. The risk rating ‘Substance abuse proclivity’ had the highest correlations with the other risk ratings, ranging from .351 with ‘Probability of involuntary departure’ to .883 with ‘Illegal substance abuse’ as noted above.

Such high correlations among the risk estimates suggest that many of the predictors are interrelated in some way. A closer look at the variables provides clarification for this finding. When considering the problem responses (e.g., items) that make up the individual risk estimates, many of the items appear to reflect very similar constructs. For example, the risk rating ‘Integrity’ evaluates integrity (e.g., honesty, responsibility, reliability, etc.), which has been consistently liked to counterproductive work behavior (Ones et al., 1993), and the risk estimate ‘Job performance’, which is derived from a combination of behaviors considered problematic for purposes of employment (CWB), are intimately connected because they share the same construct, which is deviant behavior mostly in the occupational domain. Additionally, some of the variables are simply a combination of other variables. This can be seen in the relationship of both the variables ‘Illegal drug use’ and ‘Alcohol abuse concerns’ to the risk rating
‘Substance abuse proclivity’. The ‘Substance abuse proclivity’ composite variable is made up of the items from both the aforementioned risk estimates, plus one item that reads “smokes half a pack of cigarettes, or more, a day” (Roberts & Johnson, 2001).

Lastly, the risk rating ‘Probability of involuntary departure’ consists of one category (D) from five (A,B,C,D, and F) used in the validation research for the ‘poorly suited’ risk estimate. In the validation study for the ‘Poorly suited’ risk rating, applicants given a rating of A, B, or C were considered “suitable” and those receiving ratings of D and F, were considered “poorly suited”. The D-rated applicants make up the variable ‘Probability of involuntary departure’. Therefore, the ‘Probability of involuntary departure’ is simply one element of the ‘Poorly suited’ risk rating.

Information regarding the composition of the risk ratings is important. It is necessary for the clarification of the high correlations among the variables. Information about the interrelationship of the predictors is valuable for at least three obvious reasons. First, it aids in interpreting the results of the present study. Seconds, such information informs future researchers. Third, it adds to practitioner’s knowledge regarding overlap among the variables, which has implications for interpreting the results of the CPI special police and public safety selection report. All of these issues impact both the research and practice of police psychology. Now that the relationship between the predictors and the criterion as well as the relationship among the predictors has been thoroughly treated, the discussion will turn to the topic of prediction.

Several logistic regression analyses were run to examine the validity of the CPI risk estimates in predicting applicant suitability based on background investigators ratings. Logistic regression analyses were run for each risk estimates with and without
controlling for the demographic variables ethnicity, sex, and age. Hierarchical regression analyses were run on the risk estimates with and without controlling for the demographic variables. The results from the risk rating when controlling for demographic variables will be discussed first, followed by a discussion regarding any differences in results when conducting the logistic regression results without controlling for ethnicity, sex, and age.

The results from the logistic regression analysis yielded some interesting results in term of the demographic variables and the prediction of group membership. The odds ratios for the ethnicity variables indicate that both the Native American and African American applicants are more likely to not advance in the selection process. The finding that Native American and African American applicants are less likely to pass the background investigation than Asian, Hispanic, or those who identified as “other” is predictive. The only significant result amongst the ethnicity predictors was that of African American applicants. Whether the findings are significant or not, they have important legal implications. If the use of an instrument results in systematic bias for one protected group over another, it is considered to have adverse impact (Cullen & Sackett, 2003) and is unlawful in employment settings. In addition to the finding that being a member of certain ethnic groups was predictive of whether an applicant was terminated or advanced, the results suggest that the demographic variable of sex was predictive as well. The odds ratio for the sex variables indicate that female applicants are more likely to pass the background investigation than their male counterparts. Age did not appear to predict outcome. Based on the findings, the test may produce adverse impact. Although additional research would be required to substantiate the findings, if the test does indeed result in a systematic bias favoring one group over another, the question may shift from
whether or not the CPI risk estimates are useful to whether or not the measure can be used at all.

When looking at the single predictor logistic regression analyses, all of the individual predictor models, except for the risk rating ‘Alcohol use concerns’, were significant. Therefore, all of the risk ratings save ‘Alcohol use concerns’ are predictive of background investigation outcome to some degree. The possible explanation for the lack of predictive value in the ‘Alcohol use concerns’ risk rating is the same as that stated above in terms of the nonsignificant findings when looking at the correlations between the risk rating and background investigation outcome.

The three risk ratings with the highest predictive values were ‘Probability of involuntary departure’, ‘Illegal drug use’, and ‘Poorly suited’, respectively. The findings suggest that all of the risk ratings have some predictive value in identifying who passes and who fails the background investigation. The evidence that all predictors, except for ‘Alcohol use concerns’ are related to and in some way predictive of background investigation outcome (e.g., suitability determinations) means that applicants with emotional regulation difficulties (e.g., anger management problems), a history of disciplinary reports at work, a record of law violations, and those who admit to incidents of theft, are less likely to pass the background investigation than their peers who do not evidence such tendencies through behavioral reports. Lastly, the strongest predictors of whether or not an applicant will be terminated or advanced in the selection process at the background investigation stage are those related to being rated “unsuitable” based on expert psychologist’s determinations as well as an applicants self-report of past illegal drug use.
As discussed previously, the ‘Probability of involuntary departure’ and ‘Poorly suited’ risk estimates were the result of several longitudinal studies conducted by Roberts and colleagues. The ‘Poorly suited’ risk estimate was developed using a sub-sample of 22,867 cases and involved a completed psychological evaluation of each applicant that included a face-to-face structured interview, the CPI, and other tests such as the PAI, MMPI, STAXI, and PHQ, all resulting in a suitability determination derived from four different ratings: “suitable” (A, B, and C rated) or “poorly suited” (D and F rated). The ‘Probability of involuntary departure’ represents a specific group (D-rated) of applicants found highly likely to be problematic for purposes of employment and eventually fired. Without detailed information about the criteria used by the expert psychologists to make their determinations, little can be known about what aspects are predictive in the two risk estimates derived from expert decisions. One could speculate that psychologists, being practical people, rely on the old maxim that the best predictor of future behavior is past behavior. It is more than likely that the items composing the risk estimates derived from psychologist’s ratings are concerned with more verifiable past behavior such as fighting, stealing, drug use, and various other forms of law violating behavior. Tests using verifiable behavior predict differently than those using unverifiable behaviors such as deviant or irrational thoughts. Information regarding the individual items that make up the risk estimates ‘Probability of involuntary departure’ and ‘Poorly suited’ would add significantly to an understanding of what specific behaviors are most predictive of being rated unsuitable. As for the risk estimate ‘illegal drug use’, admissions of frequent or a recent use of marijuana, cocaine, or hallucinogens as well as the sale of such drugs or driving under the influence of such drugs, are the items that make up the predictor.
Therefore, an applicant reporting past illegal drug use, the sales of illegal drugs, or driving under the influence of such intoxicants represent an increased likelihood of being found unsuitable through the background investigation.

The findings from the hierarchical logistic regression analysis suggest that the ‘Probability of involuntary departure’ and ‘Poorly suited’ risk estimates were the only significant predictors and that these two risk ratings appeared to be the most important in terms of prediction relative to the other variables. The results from the hierarchical logistic regression analysis suggest that for any variable other than ‘Probability of involuntary departure’ and ‘Poorly suited’, a change from one unit to another does not reliably increase the odds of membership in the target group. Unfortunately, the findings from the hierarchical logistic regression analysis may not be reliable. Recall the discussion regarding the impact of high correlations on the results. The intercorrelations between the scales are high, meaning that an individual scale does not add much information beyond the information in the others. This makes it very difficult to for any given variable to add much to the prediction equation once the scale with the highest relationship (e.g., ‘Poorly suited’) with the criterion (e.g., passing or failing the background investigation) has been taken onto account.

Nevertheless, based on an analysis of the interrelationship between the two most valuable predictors of ‘Probability of involuntary departure’ and ‘Poorly suited’, it was determined that the variables were moderately related and that the former accounted for that which is most predictive of suitability in the latter (e.g., D-rated applicants). Therefore, the variable ‘Probability of involuntary departure’ is likely the single most valuable predictor of the eight risk estimates in terms of predicting whether or not an
individual will either pass or fail the background investigation phase of the preemployment selection process and thus be terminated early or advanced to the next phase of the selection process. The probability of an applicant belonging to the ‘do not proceed’ group increased by a factor of 1.21 for every one point increase in the predictor ‘Probability of involuntary departure’.

Regarding the remaining risk estimates, there was only a small relationship between the other predictor variables and the outcome variable of ‘proceed’ or ‘do not proceed’. A one point increase in any other predictor (e.g., Integrity, Illegal drug use, Job performance, and Anger management), accept those that were excluded from the analyses altogether (e.g., Substance abuse proclivity and Alcohol use concerns) because prediction cannot be know about those variable left out of the analyses, was equal to and in most cases less than a 1.08 factor increase in the odds of being in the ‘do not proceed’ group. The odds ratios for the remaining risk estimates indicate that a change in estimated risk for those variables does not result in a change in odds of whether or not an individual will belong to one group or the other and therefore does not have an effect on prediction.

Based on the study results, one would conclude that those individuals who score higher on the ‘Probability of involuntary departure’ and are thus considered a risk for being fired in the future, are more likely to be terminated earlier in the screening process than their lower risk rated peers. The findings seem important from a human resources (e.g., employment selection) perspective in that the CPI appears to be useful in identifying, early in the selection process, those individuals who will likely be found unsuitable after an extensive screening process (e.g., full-field background investigation). One implication of these findings is that implementing the CPI at the earliest phase of the
personnel selection process would eliminate the waste of valuable resources such as time and money that are required to examine the suitability of applicants who will most likely be found unsuitable anyway. Additionally, from a liability standpoint, if the CPI adds information about suitability beyond that which is obtained in the background investigation, it reduces the degree of uncertainty in selecting a fit candidate and thus reduces risk.

Another important aspect of the study was the analysis of the risk estimates and their predictive values when demographic variable were not controlled for. Looking at the analyses without controlling for ethnicity, sex, and age provides information about how well the CPI risk estimates predict background outcome alone. Hiring decisions based on applicant age, sex, or ethnicity is considered discriminatory and is not legal. The use of ethnicity, sex, and age as predictors of employment decisions in the present study was done for scientific purposes. For practical use, results are needed without controlling for demographic variables. The results of the logistic regression analyses did not reveal significant differences in prediction values when the CPI risk ratings were looked at without controlling for ethnicity, sex, and age.

The data was also analyzed to examine the accuracy of the risk estimates when used together to classify (e.g., predict) the applicants in their known group (e.g., terminated for cause v. COE). In terms of classification, the CPI risk estimates (after controlling for ethnicity, age, and sex) correctly identify 71 percent of those applicants who failed the background investigation. The risk estimates also correctly identified about 78 percent of those who passed. This is an improvement over the prediction from base rates. Additionally, 21 percent of those who were predicted to fail the background
investigation actually passed the investigation (e.g., false negatives, or a specificity of .79). Conversely, 30 percent of those applicants who were predicted to pass the background investigation actually failed the investigation (false positives, or sensitivity of .70). The classification table indicated that 143 of 372 applicants were identified by the test who actually would fail the background check. To get them out of the system, 38 people who would have passed the background check would be eliminated from consideration. An important question is whether the cost in losing 38 potentially suitable applicants would be worth the savings in not doing 181 background checks. The question then becomes one of costs and benefits, which is ultimately up to the agency. However, a cost-benefit analysis may be helpful in assisting the department in making the decision.

Classification analyses were evaluated for the hierarchical logistic regression prediction model without controlling for ethnicity, age, and sex. The classification analysis indicates a good match between the observed outcomes and those predicted by the model, as was the case with the analysis controlling for demographic variables. However, the overall percentage of cases correctly classified by the model was 72.3 percent, which is slightly lower than the classification accuracy for the model when controlling for the demographic variables. The classification table predicting membership for the hierarchical regression analysis without controlling for ethnicity, age, and sex reveals that 173 individuals of the 372 applicants were correctly identified by the test who actually would fail the background check. To get them out of the system, 39 individuals who would have passed the background check would be eliminated from consideration.
Classification tables were also reviewed for the single predictor models. Based on the sensitivity and specificity values as well as the positive and negative predictive values for the classification tables, it appears that the prediction models for the individual CPI risk estimates without controlling for ethnicity, age, and sex are not as good at predicting outcome (e.g., group membership) as the prediction models for the CPI risk estimates when those demographic variables were controlled for. Additionally, similar to the single predictor models controlling for demographic variables, the prediction models without controlling for the demographic variables are better at predicting those who will fail the background check than those who will pass. Interestingly, the CPI risk rating ‘Alcohol use concerns’ demonstrated no improvement whatsoever over the prediction from base rates. The possible explanation for the failure to find significant results in terms of relationship and prediction for the ‘Alcohol use concerns’ risk estimate was described above.

Overall, the CPI risk estimates appear to be slightly more accurate at successfully identifying those individuals who are considered by the background investigators as unsuitable. Therefore, the CPI risk estimates seem to be a little more helpful in “screening out” those applicants with biographical information (e.g., backgrounds) that are considered problematic for purposes of employment. The results from the classification analyses are consistent with and provide additional support for past research and practice suggesting that the screen-out approach is the most effective strategy to date for the successful preemployment screening of police officer applicants. One important finding was that the prediction models were better at accurately classifying applicants in the known outcome groups, when demographic variables were
controlled. In terms of classification, obviously future research should examine whether the cut scores would have adverse impact and result in the undesirable loss of good applicants.

Limitations of the Present Study

As with all scientific investigations, the current study has several weaknesses that limit both the accuracy and usefulness of the results. The most important weaknesses and subsequent limitations to the current study are discussed here. By addressing the weaknesses of the current study, the hope is to provide information for future research.

The first issue that should be addressed is related to the size of the sample. The size of the sample was narrowly acceptable for the current study. The literature on applied statistics suggests having at least 50 cases for every predictor when conducting logistic regression analysis. There are currently eight Risk Estimates on the CPI Police and Public Safety Selection Report. Logistic regression analysis examining all risk estimates would not have been possible. In the present study, the hierarchical logistic regression model was conducted with six predictors. The sample size was barely sufficient. Smaller sample sizes affect the power of the study results. Based on the relatively small sample size, the results were likely not as pronounced as they would have been if the sample was much larger. Most of the more recent predictive validity studies on personality measures and police officer suitability or performance are larger. It is possible that more significant results would have emerged had the sample been larger. Future research should focus on obtaining large enough samples sizes that will accurately reflect the value of the findings.
One of the fundamental limitations in predictive validity research on police selection is that of range restriction. The present study was designed to avoid the ubiquitous threat to internal validity of range restriction. According to Ben-Porath (2003), there are very few instances where an applicant with disqualifying psychological screening results was hired by an agency. Therefore, research on the job performance of applicant found unsuitable has not been conducted. The majority of the research examining the subsequent job performance of qualified police applicants “obviously suffers from a restricted range in predictor variables which substantially reduces the likelihood of obtaining significant predictor-criterion relationships” (Hargrave et al., 1987, p. 111). The present experiment was successful in avoiding range restriction in the traditional sense. The measure under investigation in the present study was not used to make decisions about whether or not the applicants were suitable for purposes of employment. However, there are several other ways in which the range of a sample can be restricted. One way in which range restriction can enter a study is through the relationship between the predictor and any other procedures used to refine the sample (Ben-Porath, 2007). The civil service test could contribute to range restriction in as much as passing or not passing the test might be related in some way to the CPI. No method beyond the design of the study at the outset was employed for controlling the ever-present threat of range restriction. Future research might take advantage of the statistical correction technique that Ben-Porath (2007) reported.

Another issue that threatens the validity of the results in the present study has to do with confounding variables. In the case of the present study, the relationship between the predictor and the outcome variable is a real concern. One of the instruments used by
the background investigators in the determination of suitability was the PHQ. The degree
to which background investigators relied on information gather from the PHQ cannot be
known. The PHQ may have been used as a primary source of information. The risk
estimates of the CPI Police and Public Safety Selection Report were developed using
items from the PHQ as criterion. The intimate relationship between the PHQ and the CPI
risk estimates make interpreting the results very difficult. The current study could be
viewed as more of a concurrent validity study between CPI and PHQ than a study
investigating the validity of the CPI special police report in predicting police officer
suitability.

The next concern is closely related to the previously addressed weakness. The
limitation has to do with a lack of information regarding the outcome variable. The
applicants were determined suitable or unsuitable based on several job dimensions. The
job dimensions include aspects such as professional communication skills, problem
solving skills, interpersonal skills, etc. The sources of the data used for detecting problem
behaviors under the different job dimensions are numerous and varied. As mentioned
above, suitability determinations were also made using information from the PHQ.
Information about why the individuals were terminated or advanced was not included in
the present study. More specific information about what factors were used in terminating
or advancing an individual would aid in the interpretation of the results. Presently, only
gross speculation regarding the reason for termination or advancement can be made. The
lack of information regarding the investigators decision is a significant weakness in the
current study. Knowledge regarding the sources of data relied on for making each
determination would not only add to the richness of the current study, but may help in resolving the previously mentioned problem of confounding variables.

Another limitation to the present study was the interrelationship between the predictor variables. Most of the CPI risk estimates were found to be highly correlated with one another. High correlations between predictor variables can negatively impact the accuracy of the results. The existence of multicollinearity among the predictors raises a question about the reliability of the results. However, all analyses consistently reveal the importance of both the ‘Probability of involuntary departure’ and ‘Poorly suited’ variables, respectively. As mentioned previously, it is likely that the most important variable is ‘Probability of involuntary departure’, and that this variable includes that which is most predictive in the ‘Poorly suited’ risk rating. Nevertheless, in the case of the present study, the existence of multicollinearity may have rendered the results of the second analysis (e.g., hierarchical logistic regression analysis) invalid for the most part.

Further examination of the individual items (e.g. test questions) and scales that constitute the different risk ratings may help to clarify where these variables show significant overlap and may help to inform future researchers of what variable to include in their analyses and how the variables may be reconfigured to produce cleaner constructs and more useful results. The issue of overlap is no new concern when working with the CPI. Many critics of the CPI point to the overlap between scales as a major weakness of the test claiming that it interferes with statistical power. According to Groth-Marnat (2003), Gough argues that the interrelatedness of personality traits reflects the true complex nature of the human character. Gough (1987) refused to sacrifice the
accuracy of the measure by creating arbitrary boundaries between the elements of personality for the purpose of statistical eloquence.

Another weakness that requires attention involves the outcome measure. More specifically, the weakness has to do with the reliability of the suitability determinations. The reliability of interviewer ratings as criterion variables is one limitation cited in the literature (James et al, 1984). The main concern in terms of using interviewer ratings as a suitability outcome measure has to do with the subjective and qualitative nature of such ratings and the subsequent lack of reliability not only between raters but across time as well. The results of studies using subjective performance measures, despite the scientific rigor in design methodology, are “plagued” by rating errors. (Wright et al, 1990).

The concern about the reliability of interviewer ratings translates directly to a concern regarding investigator ratings because of the subjective nature of such ratings. There was no formal mechanism in place to ensure interrater reliability. Therefore, interrater reliability for background investigators was not obtained. The background investigators are provided with a detailed manual instructing them on the various behavioral components of each job dimension for which they provide ratings. The background investigation manual also instructs the investigators on where to obtain the information for each job dimension. Despite the structured nature of the background investigation, there is no insurance that each investigator makes the determination in the same way, or that the protocol is adhered to over the course of time. The latter concern is referred to as “drift” and has to do with fidelity to the protocol. It is possible that an applicant who was terminated for cause by one investigator might be advanced by another. It is also possible that one evaluator will score two similar applicants differently.
The last concern has to due with external validity. Hargrave et al. (1986) pointed out that duties performed by officers can vary widely from one agency to another and that different agency value and often require different types of individuals. The heterogeneity of police agencies in the United States and the requisite idiosyncratic job requirements result in predictable differences between successful and unsuccessful officers as a function of agency type. The implications are that research conducted on a specific type of agency may not be applicable to an agency that is significantly different.

The findings from studies examining the personality traits associated with successful applicants in a large metropolitan agency such as the NYPD or the LAPD may not apply to police officer applicants in a more rural setting. However, according to Hargrave et al. (1986), agency specific findings inform professionals who conduct psychological screening of the attributes associated with low and high performing officers making them more aware of the relevant variables and enhancing the effectiveness of selection decisions for the specific department. Therefore, agency specific findings can be more helpful than general findings. Nevertheless, the more dissimilar the setting from that which the findings were based, the less applicable the results are. The central issue is that of generalizability. The current study was conducted on a sample of police officers from a large metropolitan law enforcement agency. Therefore, the findings may not generalize to other smaller or more rural settings.

When sufficient empirical evidence is lacking to justify the use of a selection procedure (e.g., personality measure) in a situation different from that which it was validated, one of several ways to establish validity is by critically examining the technical soundness of the original study and determining the relevance of the findings to the new
situation. According to the Society for Industrial and Organizational Psychology, Inc. (2003), this specific process of generalizing validity evidence is referred to as “demonstrating the transportability of validity evidence for the selection procedure” (p. 27). Important elements in establishing whether or not validity evidence can be transported from one setting to another involve considering the comparability of job requirements, as well as the similarity of job context and candidate group (Society for Industrial and Organizational Psychology, Inc., 2003). The importance of the transportability strategy as a way to generalize validity evidence applies in those situations where local validation for a given procedure has not occurred or is not viable and a procedure such as a psychological test is thought to be useful in making selection decisions.

Conclusion

The present study was designed to investigate the validity of the CPI Suitability Risk Levels in predicting police applicant suitability as determined by background investigators. Psychological tests and biographical data have been shown to be somewhat predictive of police performance (Sarchoine, et al, 1998; Carlson, 1975; Baehr et al., 1973; Levy, 1973). The criterion for the current study was background investigators suitability determination. The determinations were based on life history indices (e.g., biographical data) such as educational background, employment history, criminal record, driving record, credit history, etc. The predictor was CPI Risk Levels, which represent applicant admissions of certain behaviors considered problematic for purposes of
employment and subsequent numerical probabilities estimating the likelihood of an applicant engaging those counterproductive behaviors.

Most studies aiming to assess selection and prediction yield unimpressive results (Snibbe and Snibbe, 1973). The results of the present study indicate that the CPI Police and Public Safety Selection Report is indeed predictive of police officer suitability. Further, the CPI special report was more successful in identifying those applicants that were found unsuitable by background investigators. Therefore, the CPI special report would function better as a screening-out technique. The screen-out approach is the standard practice in police selection presently. Therefore, the CPI appears to work in a way that is consistent with current practice.

Despite the results indicating that the CPI can provide information about the suitability of police officer applicants, the findings from the present study, in terms of goodness-of-fit, were not statistically significant; that is, the CPI Police and Public Safety Selection Report Risk Levels are considered inadequate as predictors from a hypothesis testing perspective. However, researchers often read too much into significance, or the lack thereof, and the results of inferential statistics. In fact, according to Inwald (1988), “published prediction accuracies (on the basis of classification tables presented by several researchers using different test batteries, performance criteria, and candidate groups) appear to hover around the 60 % to 75% level at best” (p. 2). Meloy (2008) also observed that reported prediction accuracies tend to hover around 75%, which is about 25 percent above chance.

No screening instrument alone can predict who will be successful and who will not (McDonough & Monahan, 1975). According to Ainsworth (1995), psychological
tests will not likely ever be independently sufficient for hiring decisions. However, such
tests can provide objective information about applicants that can be useful in determining
whether or not they are a good fit for the complex job of police officer. Barnabas (1948)
proposed that if an instrument performs better than chance, it has some utility. According
to Hibler and Kurke (1995), “selection is a dynamic process that seeks to define what
matters, and then use multiple, overlapping methods to assess the factors related to the
criterion…” (p. 61). Most selection batteries are made up of multiple assessment
measures. The CPI was found to be predictive of police applicant job suitability
determinations by background investigators. Since the current study was not looking at
actual behavior as an outcome, but rather, expert opinions of applicant suitability, it
remains unknown whether or not the CPI risk estimates predict actual on-the-job
problems (e.g., counterproductive work behavior). However, applicant records were used
in the determination of suitability, and problem behaviors detected in the background
investigation serve as the criterion for disqualification. Therefore, the CPI is likely an
adequate predictor of whether or not problem behaviors will be detected in the candidates
past.

Based on the results of the present study, the information provided by the CPI
special police report can add valuable information to the selection process, reducing
uncertainty and subsequent risk in selection. According to Bartol and Bartol (2004), one
of the primary reasons for using any screening instrument is to identify potential dropouts
or failures as soon as possible in order to save both time and money. In most cases, the
issue of whether or not to include a particular screening tool, when it is not considered
legal necessity, is a cost-benefit issue. However, because the CPI is a test of normal
personality and considered non-medical in nature, it provides the means for obtaining information about an applicant’s character that could not otherwise be gathered because the typical personality test is meant to measure pathology and is prohibited prior to a COE. Thus, without the CPI, such critically important information cannot be known before significant resources are spent. A cost analysis on the effectiveness of law enforcement selection conducted by Fitzsimons (1986) suggested that each hiring error (e.g., hiring an officer that is unsuitable) costs municipal police departments around a half million dollars. It would seem that the relatively insignificant additional cost, which is approximately $15 per applicant for test booklet, answer sheet, and scoring, of adding the CPI to the selection process would be well worth the reduction in risk. Thus, implementing the CPI as one part of the selection process does not only reduce the risk of hiring an undesirable candidate, but screens out unsuitable applicants sooner than later, saving valuable resources.
References


predictors and specific patterns of exceptional and marginal performance. The University of Chicago, Industrial Relations Center.


Enright, B. P. (2004). Personality measurement in the prediction of positive and negative


Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor and Department of Treasury. (1979). Adoption of questions and answers to clarify and provide a common interpretation of the Uniform Guidelines on Employee Selection Procedures.


Fitzgerald, P. R. (1986). The prediction of police performance using the MMPI and CPI. Dissertation Abstract International, 47, 3519. (UMI No. 862877)


test profiles for traffic officers and deputy sheriffs. *Journal of police science and administration, 14*, 250-258.


Jones v. New York City Human Resources Administration 528 F. 2d 696 (2d cir. 1976).


McCabe v Hoberman, 33 A.D. 2D 547 (1ST Dept. 1969).


Peluso v. Gourdine, NY County Supreme Court. Secial Term: Part 1, April 9, 1982. (Article 78 Proceeding).


Appendix A
Personnel Division Background Investigation SOP: 2001 Investigator Manual

Job Dimension #1 COMMUNICATIONS:
Ability to express oneself clearly in writing and speech. Ability to write a report which accurately describes what has happened. Ability to speak clearly and make oneself understood.

- SPH (sections, supplemental sheet & essay)
- Character reference questionnaires
- Educational transcripts
- PHQ interview
- SPH interview

Job Dimension #2 SKILL IN ASSESSING SITUATIONS, RECOGNIZING WHEN TO TAKE ACTION AND DECIDING ON AN APPROPRIATE COURSE OF ACTION:
Knowing how to size up a situation, identify the problem, and make a logical decision. Knowing when to take action and what kind of action is appropriate. Using good judgment in making decisions. Ability to see similarities and differences between the many situations confronted on a daily basis. Ability to consider varied and imagined solutions from both traditional and non-traditional sources.

- Character reference questionnaires
- Employment reference questionnaires
- Police/campus security contacts
- Driving record
- Section #16 (parking tickets, garnishments, delinquent taxes, et cetera)
- Credit
- Reaction to everyday challenges
- Pattern of poor judgment evidenced in several areas

Job Dimension #3 SKILLS IN READING, COMPREHENDING, RETAINING AND APPLYING WRITTEN FACTUAL INFORMATION:
Capable of learning and applying the factual material which is required of a law enforcement officer. Ability to recall factual information pertaining to laws, statutes, codes, criminal information, etc. Ability to learn and apply what is learned.

- SPH (completeness, accuracy, adherence to written and oral instructions)
- Education transcripts (pattern of poor performance, academic probation, suspension)
- Ability to follow employment rules and regulations
- Pattern of conduct suggesting a failure to learn from past mistakes (e.g., driving, credit, etc.)
Job Dimension #4 SKILLS IN ANALYZING SITUATIONS QUICKLY AND OBJECTIVELY, TO RECOGNIZE ACTUAL AND POTENTIAL DANGERS AND TO DETERMINE PROPER COURSE OF ACTION:
Applying good common sense in dealing with pressure situations. Capability of making sound decisions on the spot. Using good judgment in dealing with a potentially explosive situation. Ability to make effective, logical decisions under pressure.
- Character reference
- Employment reference questionnaires
- Driving (record, reported conduct, etc.)

Job Dimension #6 WILLINGNESS TO CONFRONT A VARIETY OF PROBLEMS AND SITUATIONS:
Possess the willingness to assertively confront and deal with a wide range of undesirable situations and events, to approach people who are behaving in a suspicious manner, to question their actions and to react positively upon discovered facts. Ability to confront a potentially dangerous situation without recklessly endangering others.
- Refer to Job Dimension #4

Job Dimension #7 INTERPERSONAL/PUBLIC RELATIONS SKILLS:
Possess the skills to establish rapport and an effective working relationship with fellow employees and with citizens from varied racial, ethnic and economic background; both individually and groups. Skills in relating with the public in a fair, tactful and courteous manner. Skills in dealing effectively with persons in various emotional states.
- Character reference questionnaires
- Employment reference questionnaires
- Police/campus security contacts
- Interactions with investigator
- Section # 16 (e.g., lawsuits, stocking orders, civil rights investigations, etc.)

Job Dimension #8 SKILLS IN OPERATING A MOTOR VEHICLE:
Qualified to possess a valid driver’s license. Skill in driving safely, under control at high speeds and in all types of weather and conditions.
- Character reference questionnaires
- Driving record
- Accident history
- Employment reference questionnaires
- Insurance claims history
- Section 16 (e.g., law suits, licenses refused, suspended, revoked, etc.)

Job Dimension #9 DEPENDABILITY AND SOUND WORK HABITS:
Have developed habits such as reporting to work on time, completing assignments on time, not malingering or abusing sick leave or other employee benefits. Following organization rules and procedures and having an acceptable discipline history.
- Employee reference questionnaires
- Employer records
Job Dimension #10 STANDARDS OF CONDUCT:
Possess the characteristics necessary to maintain credibility and to adhere to the ethical requirements of law enforcement, including:

**INTEGRITY**-refusing to yield to the temptation of bribes, gratuities, payoffs, etc., and refusing to tolerate unethical or illegal conduct on the part other law enforcement personnel. Keeping whatever is seen or hear of a confidential nature secret unless revelation is necessary in the performance of duty.

**TRUTHFULNESS**-demonstrating honesty in giving testimony or in rendering an official report or in giving any official statement about any action taken that relates to employment as a Community Police Officer.

**CREDIBILITY**-able to give testimony in a court of law without being subject to impeachment due to a reputation or history of dishonesty or due to prior criminal involvement.
Appendix B

The material (PHQ items) is copyrighted and not included.