Prediction of Serious Misconduct in Law Enforcement Officers

Alec T. Twibell,[1] Robert E. McGrath,[1] and Matthew Guller[2]

[1]School of Psychology and Counseling, Fairleigh Dickinson University

[2]Institute for Forensic Psychology

Author Note

Robert E. McGrath https://orcid.org/0000-0002-2589-5088

Correspondence concerning this article should be addressed to Matthew Guller, Institute for Forensic Psychology, Suite 2-5 Fir Court, Oakland, New Jersey 07436, U.S.A. Email: coptest@aol.com

Disclosures:

This study was facilitated by IFP Test Services, Inc., which holds the copyright for the COPS-R, one of the measures used in the study.

Author Contribution Statement:

Alec T. Twibell was the lead contributor to formal analysis and writing, as well as an equal contributor to methodology.

Robert E. McGrath was an equal contributor to conceptualization. He also provided a supporting role in formal analysis and writing the original draft.

Matthew Guller was the lead contributor to data curation and providing resources for this study. He was an equal contributor to conceptualization and formal analysis and provided a supporting role in methodology and writing.

At this time, data, materials, and analysis code from this study are not available to the public. This study was not preregistered and was exempt from IRB approval as it used only archival data.

Abstract

This study compared 143 law enforcement officers with a history of serious misconduct and 429 matched comparators who had completed the Candidate and Officer Personnel Survey-Revised (COPS-R), an inventory developed specifically for psychological evaluation of public safety candidates. Most had also completed the Personality Assessment Inventory (PAI), which allowed for scoring of scales and indices from the PAI Police and Public Safety Report (PAI-PPSR) as well as the recently developed PAI-Plus. Correlations and *t* tests indicated significant small group differences on seven COPS-R scales and one PAI-Plus index. Analyses were repeated for officers with substance-use related problems compared with non-problem officers. These identified four COPS-R and two PAI-PPSR scales that were significant predictors. Relative risk ratio analyses indicated the clinical utility of these scales at certain cutoff scores in predicting serious misconduct in law enforcement officers and candidates.

Keywords: COPS-R, PAI, police, law enforcement, psychological assessment

Public Significance Statement:

Results of this study indicated that certain scales on the Candidate and Officer Personnel Survey-Revised (COPS-R) and the Personality Assessment Inventory (PAI) were able to predict serious misconduct in law enforcement officer candidates. These findings will help to guide interpretation of test scores in law enforcement pre-hiring psychological screenings.

Prediction of Serious Misconduct in Law Enforcement Officers

The psychological screening of law enforcement candidates during the hiring process has become standard practice in most agencies throughout the United States (Bonacum & McCrerey, 1985; California Commission on Peace Officer Standards and Training [POST], 2014). These evaluations typically include an assessment of a variety of traits that can impair a law enforcement officer's ability to perform requisite duties. Examples of potential targets for evaluation include teamwork, adaptability/flexibility, and impulse control (POST, 2014).

Varela (2000) identified the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942) and its revised version (MMPI-2; Butcher et al., 2001) as the most commonly studied measures in these evaluations. A feature the MMPI-2 shares with several other instruments often used to test law enforcement candidates, such as the Personality Assessment Inventory (PAI; Morey 2007) and the California Personality Inventory (CPI; Gough, 1987), was its original development to measure a broad range of psychological constructs in a more general population than law enforcement candidates. The first two inventories also focus primarily on psychopathology variables, though POST (2014) identified a number of personality variables that can be useful in the evaluation of law enforcement candidates. Several measures are also available that were developed specifically for use in this population, and therefore focus more specifically on constructs likely to predict law enforcement officer performance (Super, 2006). Historically the best known of these has been the Inwald Personality Inventory (IPI; Inwald et al., 1983).

Previous research has supported the utility of several categories of personality and psychopathology variables in the prediction of law enforcement officer concerns. Related to somatic issues, higher scores on the IPI Sexual Concerns and Illness Concerns scales have been

shown to predict citizen complaints against officers as well as poor supervisor ratings (Scogin et al., 1995). In contrast, scores on the Somatic Complaints scale of the PAI correlated -.25 with supervisor ratings of integrity problems (Lowmaster & Morey, 2012).

Scales reflecting emotional distress have also been found useful. Tarescavage, Corey, et al. (2015) reported that the Low Positive Emotions Scale of the MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2011) was related to poor supervisor ratings in several areas such as stress tolerance, judgment, assertiveness, and teamwork, $r = .18$ to .24, while the Emotional/Internalizing Dysfunction scale was correlated with failure to accept feedback, problems with assertiveness and control under stress, and problems in teamwork, $r = .17$ to .20. Similarly, Scogin and colleagues (1995) found higher rates of citizen complaints against officers scoring higher on the IPI Anxiety scale as well as the MMPI Psychasthenia scale. In contrast, Lowmaster and Morey (2012) found that officers with higher scores on the PAI Anxiety and Depression scales engaged in fewer acts of misconduct, $r = -.21$ to -.22.

On scales reflecting problems in thinking, the Rigidity, Undue Suspiciousness, and Unusual Experiences scales of the IPI were all shown to predict citizen complaints against officers (Scogin et al., 1995). The MMPI-2-RF Ideas of Persecution and Aberrant Experiences scales have also been correlated with use of excessive force, deceptiveness, and failure to cooperate with supervisors, $r = .16$ to .23 (Sellbom et al., 2007; Tarescavage, Fischler, et al., 2015).

In terms of behavioral difficulties, the Antisocial Behavior, Aggression, and Aggressiveness scales of the MMPI-2-RF and the Antisocial Features scale of the PAI have been significantly correlated with outcomes such as citizen complaints, internal affairs complaints, insubordination, poor citizen relations, and neglect of duty, $r = .08$ to .36 (Roberts et al., 2018;

Tarescavage, Fischler, et al., 2015; Weiss et al., 2005). The CPI Self-Control and Responsibility

scales have also been shown to distinguish problem officers from non-problem officers, $d = .40$

to .60 (Hargrave & Hiatt, 1989; Sarchione et al., 1998).

Several scales related to interpersonal difficulties have also shown correlations with

problematic officer performance, including restraint and control problems and failure to accept

feedback, among others. These scales include the Social Avoidance, Interpersonal Passivity, and

Family Problems scales of the MMPI-2-RF, $r = .18-.23$ (Roberts et al., 2018; Tarescavage,

Corey, et al., 2015), as well as the Socialization scale of the CPI, $d = .29-.47$ (Hargrave & Hiatt,

1989; Sarchione et al., 1998). Lowmaster and Morey (2012) found officers scoring higher on the

Negative Relationships scale of the PAI were rated more poorly on job performance by

supervisors, $r = -.23$. Scogin and colleagues (1995) found the MMPI Cynicism scale was related

to citizen complaints against officers.

Concerns about impression management caused by the high-stakes nature of hiring

evaluations (Ones & Dilchert, 2004) have also led researchers to look for significant

relationships between response bias indicators and job performance. Past research has

demonstrated a tendency to generate elevated scores on positive impression management scales

(Lowmaster & Morey, 2012; Tarescavage, Corey, et al., 2015). Lowmaster and Morey (2012)

found that the ability of many PAI scales to predict officer problems was greater in applicants

with lower scores on the Positive Impression Management scale. Additionally, high scores on the

MMPI-2 L scale have predicted problems such as inappropriate weapon use, citizen complaints,

and property damage (Weiss et al., 2013).

Clearly, there is reason to believe a variety of personality and psychopathology variables

have the potential to predict problems functioning in the role of a law enforcement officer.

Instruments that address a broad spectrum of such variables have particular potential as predictors of job performance in this population.

The Candidate and Officer Personnel Survey-Revised (COPS-R; McGrath & Guller, 2014-2020) and the PAI Police and Public Safety Report (PAI-PPSR; Roberts et al., 2019) are two relatively recent tools that have been developed specifically for purposes of evaluating public safety candidates. The COPS-R consists of 240 true-false items reflecting personality and psychopathology variables as well as biographical data. Items and scales were initially developed in 1988 using a rational-intuitive strategy. Items were then revised over the ensuing years as clinical experience with the instrument grew. A more formal revision using empirical methods began in 2014, the result of which was the COPS-R. The current version includes nine Risk-Level scales, four Positive Descriptor scales, 16 Negative Descriptor scales, and two Potential Validity Concerns scales.

The Risk-Level scales were initially developed using a criterion-keyed approach, based on supervisor ratings collected several years after the officer completed an earlier version of the instrument. These item sets were then refined using analyses focusing on item homogeneity. The Positive Descriptor scales were part of the original instrument, and consisted of items that clinical experience had suggested were useful in identifying successful candidates. The 16 Negative Descriptor scales include 11 scales from the original instrument that were considered useful in the detection of problematic officers and five newer item sets based on latent structural analyses of the item pool as a whole. Finally, the two Potential Validity Concerns scales are intended to detect non-content responding and positive impression management.

The PAI-PPSR is based on the 22 standard scales of the PAI, as well as seven additional PAI indices specifically developed for public safety officers (police officers and firefighters).

The Poorly Suited index was developed in a study comparing candidates rated as unacceptable versus acceptable for the position by psychologists with expertise in public safety screening (Roberts et al., 2019). The other six indices—Job Performance, Integrity, Anger Management, Alcohol Use Concerns, Illegal Drug Use, and Substance Abuse Proclivity—were developed using a sample of 37,700 public safety applicants who completed the PAI. Items were chosen and cross-validated based on whether they were associated with a prior or current history of personal problems likely to interfere with their roles. These histories were self-reported at the time of testing. The COPS-R and PAI-PPSR score reports also include critical items specific to potential public safety job performance problems.

Several weaknesses are evident in the protocols used to develop the PAI-PPSR and COPS-R. The PAI-PPSR was developed using historical behavior rather than behavior once employed. Clinical experience rather than actual job performance was used to identify items for the PAI-PPSR Poorly Suited index as well as for a number of the COPS-R scales. Despite these limitations, both the COPS-R and PAI-PPSR are in frequent use. The former is administered approximately 6,000 times per year (M. Guller, personal communication, December 21, 2020), the latter approximately 25,000-30,000 times each year (R. Roberts, personal communication, December 21, 2020). Furthermore, neither has been extensively studied. One study has been published using an earlier version of the COPS-R (McGrath & Guller, 2009). It found acceptable internal reliability for scores on the COPS-R scales, and supported the instrument's ability to distinguish between officers referred for promotion and those referred for fitness for duty evaluations due to misconduct or mental health concerns (mean $r = .49$). To date, no studies have been published using the PAI-PPSR. Given there are grounds to question the validity of at least

some scales on both instruments, combined with their popularity, evaluation of their validity for the detection of problematic law enforcement candidates is critical.

The present study incorporated two beneficial methodological elements. First, it included departments of various sizes, as detailed in Table 1. These results may therefore be more generalizable than a number of prior studies that focused exclusively on large police departments (e.g. Lowmaster & Morey, 2012; Roberts et al., 2018; Sellbom et al., 2007). While the latter approach makes data collection more efficient, the results may not be representative of the 73% of U.S. police departments with fewer than 25 officers (Banks et al., 2016).

Second, the present study used only objective sources of evidence to identify officers who had committed counter-productive work behaviors (CWBs) or serious misconduct. POST (2014) defined CWBs as behaviors such as illegal drug use, violence, harassment, and destruction of property, among others. Sources for the present study included, for example, public records of assault or driving while intoxicated. In contrast, previous studies have tended to rely on subjective supervisor ratings as outcome measures (e.g. Lowmaster &. Morey, 2012; Roberts et al., 2018), so that results were potentially influenced by memory, personal relationships between supervisors and officers, and variations in supervisor judgment. McGrath and Guller (2009) found that prediction of supervisor ratings was attenuated when interdepartmental variation was controlled for, suggesting that these variations are largely substantive rather than artifactual, but most existing studies have not evaluated this issue.

Method

Participants

The target group of problem cases included 143 law enforcement officers (police and other similar non-correctional peace officers) who had received a pre-employment psychological

evaluation through the Institute for Forensic Psychology (IFP), a psychological assessment agency, in New Jersey or New York between 2008 and 2020 for which concerns were subsequently raised about serious misconduct. Officers were included in this group for one of five reasons:

- They were later referred by the department back to the IFP for a fitness for duty evaluation, i.e., a psychological evaluation of a serving officer. Referred officers were only included in this study as a problem case if the reason for referral resulted in disciplinary action against the officer, or if during the course of the evaluation it was learned the officer had been disciplined at some earlier point in their career.

- Department personnel contacted IFP about the officer either because they were experiencing problems with them as a recent hire, or they were considering requesting a fitness for duty evaluation. The latter cases were distinct from the first group in that the re-evaluation was not requested as the problem was deemed to be disciplinary, or the officer resigned or was terminated prior to the evaluation.

- New Jersey and New York newspapers were reviewed online by research staff for articles indicating an officer had been found guilty of serious misconduct.

- Internet search of public civil service records indicated suspension or termination.

- The individual returned for another pre-employment evaluation in which it was determined that he or she had been disciplined in or terminated from their previous position.

Misconduct ranged in seriousness from cheating on an examination at the police academy to conviction for manslaughter. Because candidate assessment is intended to identify potential for serious fitness concerns at any time subsequent to hire, no time limit was placed on the time interval between testing and occurrence of the problematic incident. Problems developed

anywhere from several months after testing (while in the training academy) to 11 years later. Unfortunately, the interval from testing to incident was not recorded. The number of problem cases identified via each source as well as demographics for the sample as a whole are provided in Table 2.

The comparison group consisted of 429 matched comparators, three per problem case, who had also participated in pre-employment evaluation. Each set of three was matched with one problem officer on the following criteria: they had applied for a position with the same job title; they had been recommended for hiring; to the extent possible, they were employed by the same agency that hired the problem case (296 out of 429); if matching by agency was not possible, matches were drawn from agencies that were similar to that of the problem case in size, community type (city, suburban, or rural), and use of a civil service process in hiring. In terms of demographics, cases were matched exactly on race and gender, and where possible, approximate equivalence in level of education (mean difference = 0.5) and age (mean difference = 3.9 years).

Problem officers came from 67 different agencies. In 86 of the 143 cases (60.1%), problem officers were able to be matched with three officers from their own department. In 13 cases (9%), problem officers had two matches from their own agency and one from another agency. In 17 cases (12.0%), one match was found within the problem officer's own agency and two were used from outside agencies. In the remaining 25 cases (17.5%), all matches came from an agency other than the one the problem officer was with. Matching was often complicated by factors including small agency size and difficulty matching demographic information that was deemed important in the comparisons, especially ethnicity and years of education. Of the control group, there were 32 agencies from which matches were included that did not have any problem

officers included in the study. Of the problem officer agencies, 9 were not represented among the control cases. Thus, problem and control cases were pulled from a total of 90 agencies.

Table 2 provides information on broad categories of misconduct problem officers engaged in. Table S1 in the Supplementary Materials provides more comprehensive data for the nature of misconduct problem officers engaged in. Of the 143 problem officers, 52 were formally disciplined and 65 were terminated. Data on consequences were unavailable for 16 officers who were convicted or plead guilty to driving while intoxicated and 10 who entered substance abuse rehabilitation programs. These 26 participants make up the majority of the 38 problem officers disciplined or terminated for substance-related problems, while four others were arrested for possession of illicit drugs. Temporary restraining orders were issued in one assault case and three domestic violence cases. Assaults occurred both on- and off-duty. Beyond the major categories of offense outlined in Table 2, problem officers were disciplined or terminated for offenses including neglect of duty, insubordination, falsifying records, and poor attendance.

Measures

Candidate and Officer Personnel Survey-Revised

As described previously, the COPS-R consists of 240 true-false items from which 31 scales are generated. The scales were developed using a combination of strategies, with a particular focus on rational-intuitive and criterion-keyed approaches to the identification of candidate items, and refinement of item sets using inter-item relationships. This measure can be administered online or by paper-and-pencil. The latter can be scored online or using optical character recognition equipment. The COPS-R normative sample consists of 10,566 pre-hire public safety candidates (M. Guller, personal communication, September 12, 2019).

Personality Assessment Inventory

The PAI is a widely used self-report measure of psychopathology and psychological functioning consisting of 344 items answered on a four-choice scale from *False* to *Very True*. The original set of scales included four primary validity scales, 11 clinical scales, five scales reflecting treatment issues, and two reflecting interpersonal style. The publisher has recently introduced an expansion of the scale set (PAI-Plus) that provides an additional 12 indicators of validity, 14 supplementary clinical indicators, and 31 scales reflecting personality pathology. The PAI-PPSR was developed under license from the publisher of the PAI, and as described above added seven indices reflecting problems related to public safety officer performance.

Procedure

All participants were initially evaluated as part of a standard job offer psychological screening process. Because the evaluation includes questions reflecting mental health status, participants had been conditionally offered employment at the time of the evaluation in adherence to requirements of the Americans with Disabilities Act. All 572 participants had completed the COPS-R. Out of 143 problem officers, 72 (50.3%) were tested prior to 2014 and so completed the original version of the COPS; 71 (49.7%) were tested using the COPS-R. Item content was the same across the two versions except for 19 items that were minimally re-worded for the COPS-R for greater clarity or to address privacy concerns. All scores were converted to *T* scores based on the COPS-R normative sample. The comparison cases all completed testing around the same time as their matched problem officer, and in all cases completed the same version of the COPS. Because the PAI was added to the selection battery after the COPS-R, only 459 participants (115 problem officers and 344 members of the comparison group) had also completed the latter instrument. These participants initially completed the standard PAI, and

scores on PAI-PPSR and PAI-Plus scales and indices were calculated at a later time. *T*-scores were used for all analyses.

At this time, data, materials, and analysis code from this study are not available to the public. This study was not preregistered and was exempt from IRB approval as it used only archival data.

Results

Groups were compared using *t*-tests for correlation coefficients. No correction for multiple comparisons was implemented, as this is the first investigation of these scales, so we wanted to provide as inclusive an analysis of significant findings as possible. No correction for range restriction was implemented, for several reasons. First, standard deviations for the sample were not substantially lower than the normative value of 10 for *T* scores. Second, this allowed greater comparability with prior studies, most of which have reported uncorrected values. Third, the outcome variable was binary (problem versus comparison), a situation in which there is no standard method of correction procedures (Pfaffel et al., 2016).

These analyses revealed significant group differences for one PAI-Plus index as well as seven COPS-R scales (see Table 3; Table S2 in the Supplementary Materials for results on all scales and indices). In each case, problem officers scored higher on scales indicative of greater risk, and lower on COPS-R scales indicative of good job performance.

For each scale or index in Table 3, a series of diagnostic efficiency analyses were then conducted. Scores were dichotomized based on cut scores in 5-point increments from $T = 45$ to 65. Table 4 presents relative risk ratios (RRRs) for these analyses whenever the selection ratio fell between 3% and 20% (Tarescavage, Corey, et al., 2015). The selection ratio indicates the percentage of the sample that scored at or above the cut score for that analysis, with the range of

3-20% suggesting identification of a clinically useful number of cases. The RRR represents the relative risk of a problem officer versus a comparison officer obtaining a score at or above the cut score (or below the cut score for COPS-R Positive Descriptor Scales). The RRR is statistically significant at $p < .05$ if the 95% confidence interval does not overlap with 1.0. RRRs were significant for at least one cut score within the accepted selection ratio for the COPS-R Positive Descriptor Scale Self-Discipline using 60 T as the cut score, and for Negative Descriptor Scales Antisocial Activities and Impulsivity at either 55 or 60 T (see Table 4; S3 in the Supplementary materials for results at all score cutoffs).

Additional t-tests were conducted to assess differences in scale scores between participants with substance abuse related misconduct and non-problem officers (see Table 5; S4 in the Supplementary Materials for results on all scales). Officers with substance abuse problems scored significantly higher on the PAI-PPSR Alcohol Use Concerns and Substance Abuse Proclivity indices. Surprisingly, these officers scored significantly lower on the COPS-R Risk Level scale Global Prediction and Productivity, and higher on the Positive Descriptor Scale Motivation. However, substance abuse problem officers also scored higher on the Negative Descriptor Scales Antisocial Activities and Substance Abuse. RRRs both achieved a clinically useful selection ratio and were significant at one or more cutoff scores for the PAI-PPSR Substance Abuse Proclivity and Alcohol Use Concerns, as well as the COPS-R Negative Descriptor Scales Antisocial Activities and Substance Abuse (see Table 6; S5 in the Supplementary Materials for results at all score cutoffs).

Discussion

The present study sought to examine the ability of the PAI and the COPS-R to predict post-hire serious misconduct in law enforcement officer candidates. Law enforcement officers

bear the responsibility of protecting civilians, but in doing so have the potential to cause

substantial harm. Therefore, screening out candidates likely to engage in serious misconduct is

critical to ensuring safety within the community. The COPS-R and PAI-PPSR were designed

specifically for use in screening these candidates. However, this is the first study to date

examining their validity and utility in predicting officer misconduct post-hire. It seems also to be

the first to use indicators of misconduct independent of departmental records.

The results of the present study support the notion that scores on some of these scales can

predict problem officers. Based on meta-analytic evidence that correlations of .10 are small, .20

are medium-sized, and .30 are large (Gignac & Szodorai, 2016), effect sizes were medium-sized

for COPS-R Negative Descriptor scales and small for Risk Level scales, Positive Descriptor

scales, and for the PAI-Plus Back Random Responding scale. This last finding was the least

expected, and suggested that poorer candidates may have a tendency to lose focus during the

assessment. All eight scales were associated with problem officer status in the expected direction

(i.e., problem officers consistently scored higher on negative and lower on positive indicators).

The correlations for the Negative Descriptor scales Antisocial Activities and Impulsivity were

the largest in the set. Given this study's focus on severe misconduct this is unsurprising, as these

two characteristics seem likely to be important contributors to the occurrence of behaviors

warranting disciplinary action.

COPS-R scales that differed significantly in problem officers appear to relate to two

constructs in particular. The Relations with Public and Social Adjustment scales suggest poorer

interpersonal abilities in problem officers. The Impulsivity, Antisocial Activities, Self-

Discipline, and Judgment scales all relate to one's ability to refrain from engaging in problematic

behavior. The Success scale is an amalgam of items, including items relating to a lower rate of

problematic behavior (e.g., potential for substance abuse and impulsivity), and better

interpersonal skills (e.g., less bias against certain demographic groups). The Success scale

therefore appears to encompass both of the primary domains that are related to serious

misconduct in this sample.

Relative risk ratios, while simultaneously generating a clinically useful selection ratio and

significant for some scales at certain cut scores, were all < 2.0 for the sample as a whole.

However, even small effects should be considered meaningful in the prediction of serious

misconduct, because of the potential harm that may be caused by even a single incident across an

officer's entire career. It should also be noted that the prediction of problem behavior many years

after testing, which is a potentially unrealistic goal of testing with law enforcement candidates, is

likely to result in small overall effects. It should also be noted that, unlike many previous studies,

the present study included only serious misconduct, so any ability to predict these outcomes is

important. Based on these arguments, we believe these measures can play a useful role in the

comprehensive assessment of law enforcement candidates. Whether this conclusion also applies

to public safety candidates more generally unfortunately remains uncertain. Given the extent of

vetting for law enforcement candidates prior to a conditional offer of employment, though, the

potential exists for these instruments to prove more effective in the context of other positions

such as firefighter or corrections officer.

Results were similar for the more homogeneous set of officers who demonstrated

subsequent problems with substance abuse. Scales directly addressing substance abuse

tendencies were associated with risk ratios ≥ 2.25. Surprisingly, these officers scored lower on

COPS-R Risk Level scales Global Prediction and Productivity. It is unclear why these scale

scores would be lower in the substance use problem group. It may well be that other strengths

offset indicators of their abuse potential during earlier evaluation.

Limitations to this study include the method by which data were collected. While records

were searched for documented incidents of misconduct, there was no way of confirming with

certainty that "non-problem" officers had not engaged in misconduct at some point during their

careers. Therefore, the non-problem officer group is in reality a "no known problem" group.

Nonetheless, the large size of this group likely ensures that mean scores were representative of

officers without serious misconduct. A second issue is that the base rate of problem behavior,

which affects both correlational and diagnostic efficiency statistics, was artificially manipulated.

It is difficult to gauge the actual rate of serious misconduct in the population of law enforcement

officers. There is research to suggest the rate of domestic violence among police officers is about

8% (see Erwin et al., 2005). Additionally, Ballenger et al. (2016) found that roughly 11% of

male officers and 16% of female officers had recently engaged in alcohol use reaching "at-risk"

levels according to National Institute on Alcohol Abuse and Alcoholism standards. Thus, it is

likely that serious misconduct as defined in this study is relatively common among law

enforcement officers, and the base rate of 25% in this sample may not be markedly inaccurate.

Finally, this sample demonstrated some degree of range restriction. Therefore, it should be noted

that the cut scores suggested here may overestimate risk levels if applied to samples with greater

variability in scale scores. Mean scores and standard deviations in this sample differ substantially

from some previous samples (e.g. Lowmaster & Morey, 2012; Tarescavage, Corey, et al., 2015).

The variability across samples unfortunately impedes efforts to establish universal standards for

test score interpretation in this field, as clinicians in different settings are likely to see different

scores on these measures.

Overall, this study supports the presence of differences in PAI-PPSR and COPS-R scores between officers who went on to engage in serious misconduct and those who did not. Given the ease of administration, we encourage the continued use of these instruments with law enforcement candidates, and perhaps other public safety candidates as well. However, it should be noted that these measures are intended for use within a larger assessment, including interviewing and review of prior behavioral difficulties. Future research should address the integration of multiple measures in hiring and preemployment psychological assessments.

References

Ballenger, J. F., Best, S. R., Metzler, T. J., Wasserman, D. A., Mohr, D. C., Liberman, A.,

Delucchi, K., Weiss, D. S., Fagan, J. A., Waldrop, A. E. & Marmar, C. R. (2011).

Patterns and predictors of alcohol use in male and female urban police officers. *The*

American Journal on Addictions, 20(1), 21-29. https://doi.org/10.1111/j.1521-

0391.2010.00092.x

Banks, D., Hendrix, J., Hickman, M., & Kyckelhahn, T. (2016). *National sources of law*

enforcement employment data (Report No. 249681). U.S. Department of Justice, Office

of Justice Programs, Bureau of Justice Statistics.

Ben-Porath, Y. S., & Tellegen, A. (2011). *MMPI-2-RF: Manual for administration, scoring and*

interpretation. University of Minnesota Press.

Bonacum, W. T., & McCrerey, P. W. (1985). *Psychological screening and testing*. Virginia

Department of State Police.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom,W. G., & Kaemmer, B.

(2001). *MMPI–2: Manual for administration, scoring, and interpretation* (rev. ed.).

University of Minnesota Press.

California Commission on Peace Officer Standards and Training. (2014). *Peace officer*

psychological screening manual. Author.

Erwin, M. J., Gershon, R. R. M., Tiburzi, M., & Lin, S. (2005). Reports of intimate partner

violence made against police officers. *Journal of Family Violence, 20*(1), 13-19.

https://doi.org/10.1007/s10896-005-1505-3

Fischler, G. L. (2004). *Identifying psychological predictors of police officer integrity problems*.

U.S. Department of Justice, Office of Community Oriented Policing Services.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences

researchers. *Personality and Individual Differences, 102*, 74-78.

https://doi.org/10.1016/j.paid.2016.06.069

Gough, H. G. (1987). *CPI manual* (3rd ed.). Consulting Psychologists Press.

Hathaway, S. R., & McKinley J. C. (1942). *Manual for the Minnesota Multiphasic Personality*

Inventory. University of Minnesota Press.

Hargrave, G. E., & Hiatt, D. (1989). Use of California Psychological Inventory in law

enforcement officer selection. *Journal of Personality Assessment, 53*(2), 267-277.

https://doi.org/10.1207/s15327752jpa5302_5

Hiatt, D., & Hargrave, G. E. (1988). MMPI profiles of problem peace officers. *Journal of*

Personality Assessment, 52(4), 722-731. https://doi.org/10.1207/s15327752jpa5204_11

Inwald, R., Knatz, H., & Shusman, E. (1983). *Inwald Personality Inventory manual*. Hilson

Research.

Lowmaster, S. E., & Morey, L. C. (2012). Predicting law enforcement officer job performance

with the personality assessment inventory. *Journal of Personality Assessment, 94*(3),

254-261. https://doi.org/10.1080/00223891.2011.648295

McGrath, R. E., & Guller, M. (2009). Concurrent validity of the Candidate and Officer Personnel

Survey (COPS). *International Journal of Police Science & Management, 11*(2), 150–159.

https://doi.org/10.1350/ijps.2009.11.2.119

McGrath, R. E., & Guller, M. (2014). *C.O.P.S.: Candidate and Officer Personnel Survey –*

Revised Technical Manual. IFP Test Services, Inc.

Morey, L. C. (2007). *Personality Assessment Inventory (PAI): Professional manual* (2nd ed.).

Psychological Assessment Resources.

Ones, D. S., & Dilchert, S. (2004). *Personality and police officer work performance: A construct-based, comprehensive meta-analysis and implications for pre-offer screening and psychological evaluations*. Paper presented at the Annual Meeting of the International Association of Chiefs of Police, Los Angeles, CA.

Pfaffel, A., Kollmayer, M., Schober, B., & Spiel, C. (2016). A missing data approach to correct for direct and indirect range restrictions with a dichotomous criterion: A simulation study. *PLoS ONE, 11*(3), e0152330. https://doi.org/10.1371/journal.pone.0152330

Roberts, M. D., Roberts, R. M., & Johnson, M. (2019). *PAI: Police and Public Safety Report-Version 2: Technical manual*. Johnson, Roberts, and Associates. Retrieved from https://www.jra-assessment.com/jra-resources.html.

Roberts, R. M., Tarescavage, A. M., Ben-Porath, Y. S., & Roberts, M. D. (2018). Predicting postprobationary job performance of police officers using CPI and MMPI–2–RF test data obtained during preemployment psychological screening. *Journal of Personality Assessment, 101*(5), 544–555. https://doi.org/10.1080/00223891.2018.1423990

Sarchione, C. D., Cuttler, M. J., Muchinsky, P. M., & Nelson-Gray, R. O. (1998). Prediction of dysfunctional job behaviors among law enforcement officers. *Journal of Applied Psychology, 83*(6), 904–912. https://doi.org/10.1037/0021-9010.83.6.904

Scogin, F., Schumacher, J., Gardner, G., & Chaplin, W. (1995). Predictive validity of psychological testing in law enforcement settings. *Professional Psychology: Research and Practice, 26*(1), 68-71. https://doi.org/10.1037/0735-7028.26.1.68

Sellbom, M., Fischler, G. L., & Ben-Porath, Y. S. (2007). Identifying MMPI-2 predictors of police officer integrity and misconduct. *Criminal Justice and Behavior, 34*(8), 985–

1004. https://doi.org/10.1177/0093854807301224

Super, J. T. (2006). A survey of pre-employment psychological evaluation tests and procedures.

    Journal of Police and Criminal Psychology, 21(2), 83-87.

    https://doi.org10.1007/bf02855686

Tarescavage, A. M., Corey, D. M., Gupton, H. M., & Ben-Porath, Y. S. (2015). Criterion

    validity and practical utility of the Minnesota Multiphasic Personality Inventory–2–

    Restructured Form (MMPI–2–RF) in assessments of police officer candidates. *Journal of*

    Personality Assessment, 97(4), 382–394. https://doi.org/10.1080/00223891.2014.995800

Tarescavage, A. M., Fischler, G. L., Cappo, B. M., Hill, D. O., Corey, D. M., & Ben-Porath, Y.

    S. (2015). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-

    RF) predictors of police officer problem behavior and collateral self-report test scores.

    Psychological Assessment, 27(1), 125-137. https://doi.org/10.1037/pas0000041

Varela, J.G. (2000). *Meta-analysis of the predictive validity of personality testing in law*

    enforcement employment settings [Doctoral Dissertation, University of Alabama].

    ProQuest Dissertations & Theses Global: The Sciences and Engineering Collection.

    https://libaccess.fdu.edu/login?url=https://www-proquest-

    com.libaccess.fdu.edu/dissertations-theses/meta-analysis-predictive-validity-

    personality/docview/304577037/se-2?accountid=10818

Weiss, W. U., Zehner, S. N., Davis, R. D., Rostow, C., & DeCoster-Martin, E. (2005).

    Problematic police performance and the Personality Assessment Inventory. *Journal of*

    Police and Criminal Psychology, 20(1), 16-21. https://doi.org/10.1007/bf02806703

Weiss, P. A., Vivian, J. E., Weiss, W. U., Davis, R. D., & Rostow, C. D. (2013). The MMPI-2 L

    scale, reporting uncommon virtue, and predicting police performance. *Psychological*

Services, *10*(1), 123-130. https://doi.org/10.1037/a0029062

Table 1

Department Sizes.

Number of Officers	N	%
1-20	17	11.9
21-30	15	10.5
31-50	26	18.2
51-75	6	4.2
76-100	8	5.6
101-250	32	22.4
251-500	6	4.2
501-1000	19	13.3
1001-3000	14	9.8

Note. N refers to the number of problem cases coming from departments of that size range.

Table 2

Demographic Statistics.

Characteristic	*N*	*%*
Gender		
Male	520	90.9
Female	52	9.1
Primary Ethnicity		
Caucasian	340	59.4
African American/Black	128	22.4
Hispanic	79	13.8
Asian/Pacific Islander	18	3.1
Other	7	1.2
Level of Education		
High school equivalency diploma	9	1.6
High school graduate	77	13.5
Some college	158	27.6
Associate's degree	88	15.4
Bachelor's degree	228	39.9
Graduate degree	12	2.1
Problem/Misconduct Type (*N* = 143)		
Substance abuse/Alcohol related offense/DWI/Drug rehab	38	26.6
Domestic violence/Harassment/Stalking/TRO	12	8.4
Non-domestic violence assault charge, alcohol not involved	38	26.6
Disciplined or terminated for other misconduct	55	38.5

Note. Participants were between the ages of 18 and 55 (*M* = 26.9, *SD* = 5.1). DWI = driving while intoxicated; rehab = rehabilitation; TRO = temporary restraining order.

Table 3

Significant t-Test Results for All Participants.

Scale	Comparison Group		Problem Officers			
	M	*SD*	*M*	*SD*	*r*	*se*
PAI-Plus						
Back Random Responding	46.02	3.89	46.91	4.60	0.10*	0.44
COPS-R						
Risk Level Scales						
Judgment	45.52	8.58	47.75	8.49	0.11*	0.83
Relations with Public	46.44	8.10	48.13	8.58	0.09*	0.79
Positive Descriptor Scales						
Success	54.57	7.18	52.59	6.67	-0.12*	0.68
Social Adjustment	54.29	6.99	52.27	6.78	-0.13*	0.67
Self-Discipline	52.98	7.74	51.10	6.94	-0.11*	0.73
Negative Descriptor Scales						
Antisocial Activities	46.45	7.03	49.93	8.57	0.20*	0.72
Impulsivity	47.10	7.20	50.69	7.78	0.21*	0.71

*$p < .05$

PAI *n* = 115 for problem officers, 345 for comparison group.
COPS-R *n* = 143 for problem officers, 429 for comparison group.

Table 4

Relative Risk Ratios for Scale Scores and Group Membership.

Scale	Cutoff (≥)	Selection Ratio	Risk If Elevated	Risk If Not Elevated	Relative Risk Ratio	*CI* (95%)
COPS-R						
Positive Descriptor Scales						
Self-Discipline	60*T*	0.16	13.83%	27.20%	1.97*	1.16-3.33
Negative Descriptor Scales						
Antisocial Activities	60*T*	0.09	44.90%	23.14%	1.94*	1.37-2.75
Antisocial Activities	55*T*	0.15	42.35%	21.97%	1.93*	1.43-2.60
Impulsivity	60*T*	0.12	39.39%	23.12%	1.70*	1.21-2.39
Impulsivity	55*T*	0.16	37.36%	22.66%	1.65*	1.21-2.26

* $p < .05$

Note. SA = Substance Abuse.

PAI *n* = 115 for problem officers, 345 for comparison group.

COPS-R *n* = 143 for problem officers, 429 for comparison group.

Table 5

Significant t-Test Results for Substance Abuse Problem Groups.

	Comparison Group		Substance Abuse Problem Officers			
	M	*SD*	*M*	*SD*	*r*	*se*
PAI-PPSR						
Substance Abuse Proclivity	45.28	5.21	47.37	5.94	0.11*	0.96
Alcohol Use Concerns	46.67	4.44	48.71	5.39	0.13*	0.83
COPS-R						
Risk Level Scales						
Global Prediction Scale	45.43	8.29	42.31	4.48	-0.11*	1.35
Productivity	45.58	8.54	42.82	5.33	-0.09*	1.39
Negative Descriptor Scales						
Antisocial Activities	46.45	7.03	50.72	7.97	0.16*	1.19
Substance Abuse	48.01	7.71	51.10	9.23	0.11*	1.31

*p < .05
PAI *n* = 33 for problem officers, 345 for comparison group.
COPS-R *n* = 39 for problem officers, 429 for comparison group.

Table 6

Relative Risk Ratios for Scale Scores and Substance Abuse Problem Groups.

Scale	Cutoff ($\geq$)	Selection Ratio	Risk If Elevated	Risk If Not Elevated	Relative Risk Ratio	*CI* (95%)
PAI-PPSR						
SA Proclivity	55*T*	0.08	22.58%	7.49%	2.96*	1.42-6.18
Alcohol Use Concerns	60*T*	0.03	30.00%	8.15%	3.65*	1.34-9.93
Alcohol Use Concerns	55*T*	0.08	23.33%	7.47%	3.06*	1.47-6.39
COPS-R						
Negative Descriptor Scales						
Antisocial Activities	60*T*	0.07	20.59%	7.37%	2.79*	1.33-5.85
Antisocial Activities	55*T*	0.13	18.33%	6.86%	2.67*	1.41-5.08
Substance Abuse	60*T*	0.12	16.36%	7.26%	2.25*	1.13-4.49
Substance Abuse	55*T*	0.12	16.36%	7.26%	3.25*	1.13-4.50

* $p < .05$

Note. SA = Substance Abuse.

PAI *n* = 33 for problem officers, 345 for comparison group.

COPS-R *n* = 39 for problem officers, 429 for comparison group.