

A LONGITUDINAL MODEL FOR PREDICTING PERFORMANCE OF POLICE OFFICERS USING PERSONALITY AND BEHAVIORAL DATA

CARLOS G. FORERO
DAVID GALLARDO-PUJOL
ALBERTO MAYDEU-OLIVARES
ANTONIO ANDRÉS-PUEYO
University of Barcelona

A growing body of evidence suggests that the performance of police officers depends on individual dispositions. Although predictive studies of police performance using personality measures often focus on academy training, in this study the authors fitted a longitudinal structural equation model to examine the predictive validity of the training process for actual performance after graduation. To do so, the authors used behavioral and self-reported personality measures as predictors in a sample of 2,010 police candidates enrolled at the Catalan Institute for Public Safety. While academic qualifications alone predicted 27.3% of performance variance, the predictive power was greatly improved when training was included as a mediator of personal dispositions. The full final model accounted for 60% of performance variance. Results suggest that actual job performance is indeed influenced by personality, but that this influence is mediated by training.

Keywords: law enforcement personnel; personnel selection; 16PF; CAQ; performance prediction; individual differences; mediator variables

Psychological testing has been considered a useful tool for assessing law enforcement officers since the 19th century (Drees, Ones, Cullen, Spilberg, & Viswesvaran, 2003). Today, a growing body of evidence suggests that job performance depends to a large extent on features that are not entirely knowledge or skills, that is, on individual dispositions (Barrick & Mount, 2005; Conard, 2006; Ozer & Benet-Martínez, 2005). And although cognitive abilities have traditionally been used as the main measure to select police candidates, meta-analytic studies have shown the benefits of assessing personality traits as job predictors (see Aamodt, 2004a, 2004b; Barrick & Mount, 1991; Salgado, 1997; Tett, Jackson, & Rothstein, 1991).

Historically, personality assessment has been far from being the norm, perhaps because of the impact of situationist criticisms of the use of personality measures in personnel selection (Guion & Gottier, 1965; Mischel, 1968). Even though pure situationist perspectives

AUTHORS' NOTE: *This study was partially supported by Grants SEJ2006-08204/PSIC (Albert Maydeu-Olivares, principal investigator) and SEJ2005-09170-C04-01/PSIC (Santiago Redondo-Illescas, principal investigator), both from the Spanish Ministry of Education, and Project FBG-EPC 303886 (Antonio Andrés-Pueyo, principal investigator) from the Catalan Institute for Public Safety. We thank the Catalan Institute for Public Safety, particularly its director, Carlos González Zorrilla, and Maria Angels Soler for their collaboration and support during the project. We would also like to thank the anonymous reviewers and the editor for their most valuable comments during the preparation of the final version of the present article. Correspondence concerning this article should be addressed to Carlos García Forero, Departament de Personalitat, Avaluació i Tractament Psicològic, Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, 08035-Barcelona, Spain; e-mail: carlos.garcia.forero@ub.edu.*

have for the most part been abandoned (Funder, 2008), dispositional variables do not seem to have fully reestablished themselves as a key component in personnel selection. Barrick and colleagues (Barrick, Mount, & Judge, 2001) reported that no one method has been systematically used to compose a small, manageable set of personality predictors. They also note lack of clarity in the definitions of the traits being measured. However, there are also problems of a methodological nature. For instance, the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940), which is not a "pure" personality measure but a psychopathology screening tool, is often used in personnel selection, despite the fact that its clinical orientation may render it inappropriate in this setting because psychopathology rates in job applicants cannot be considered more frequent than in the general population (Costello, Schoenfeld, & Kobos, 1982). As a result, the MMPI has demonstrated weak prediction power in personnel recruitment, in particular in the context of law enforcement (Tett et al., 1991; Varela, Boccaccini, Scogin, Stump, & Caputo, 2004). It must be noted however that some recent research is more favorable to the use of the MMPI-2 (see e.g., Sellbom, Fischler, & Ben-Porath, 2007, and references cited therein).

In general, however, much of the evidence argues in favor of the use of nonclinical personality assessment methods in the selection of law enforcement personnel. Highly influential meta-analytic studies by Barrick and Mount (1991), Tett et al. (1991), and Salgado (1997) included police officers as one of their occupational groups to illustrate the validity of personality predictors for performance appraisal. In his comprehensive report, Aamodt (2004b) reviewed a great many studies using personality variables to assess different police performance outcomes, highlighting the importance of personal dispositions for predicting performance in these contexts. Other meta-analytic studies also indicate the usefulness of personality measures as screening tools in law enforcement contexts (Varela et al., 2004).

To complete the picture, we should bear in mind that the favorable reception of the Five Factor personality model (Costa & McCrae, 1992) in personality psychology has led to something of a revival in personality assessment. The Five Factor model has served as a communication tool between professionals on personality results and has renewed interest in trait assessment during training and personnel selection (Barrick & Mount, 1991; Mount & Barrick, 1995; Ozer & Benet-Martínez, 2005). In recent years, two studies have empirically explored the predictive validity of personality questionnaires with regard to police performance (Black, 2000; Detrick, Chibnall, & Luebbert, 2004). In the first study, Black (2000) found significant univariate correlations between the Neuroticism (N), Extraversion (E), and Conscientiousness (C) domains of the NEO Personality Inventory-Revised (NEO PI-R) and a global measure of performance that included academic performance, physical performance, handling firearms, driving, and other skills. Black considered the isolated effect of personality on prediction and performed a multivariate analysis with domain scores to predict global performance. The Conscientiousness domain added predictive power to the isolated measures of cognitive ability used, attaining a multiple correlation of .42 when using both sources of information. Detrick et al. (2004) extended the findings of Black in an American sample using disciplinary outcomes as an outcome criterion. First, they explored academic performance, finding positive relationships at the facet level between performance and the facet values from the Openness (O) domain and Excitement-Seeking from the Extraversion domain. These domains accounted for a significant percentage of performance variance (24% to 25%). These authors also found

more specific relations between personality and individual skills. For firearms performance, Anxiety was the facet that was the best predictor. Physical performance was predicted at the facet level by Deliberation (Conscientiousness), Fantasy (Openness), and Activity (Extraversion). With respect to Absenteeism, the model had an overall predictive accuracy of 95%; significant predictors were Self-Consciousness (N), Altruism (Agreeableness), Feelings (O), Order (C), Positive Emotions (E), and Vulnerability (N). Interestingly, Detrick and colleagues found that trainees who did not graduate from the academy scored significantly higher on Depression, Impulsiveness, and Vulnerability scores (N) and lower on Competence (C).

All these findings are in agreement with the results of meta-analyses by Barrick and Mount (1991) in the United States and by Salgado (1997) in the European Union, which identified these traits as important personality factors for police performance. These results also corroborate the existing literature on job personality and job performance in law enforcement and non-law enforcement populations (Barrick & Mount, 1991; Piedmont & Weinstein, 1998; Salgado, 1997; Tett et al., 1991).

In spite of this support for trait assessment in personnel selection, different occupational contexts demand different personality traits, and assessment tools can be explicitly aimed to appraise these traits either in isolation or in combination (Tett et al., 1991). With this in mind, some tools have been specifically used to predict police officer performance. Perhaps the most specific measurement system devoted to detect high performers among applicants is the Law Enforcement Assessment and Development Report (LEADR; IPAT, 1987; IPAT & Staff, 1998), which focuses on personality patterns. It consists of a set of composite scores based on the 16 Personality Factor Questionnaire (16PF-5; Cattell, Cattell, & Cattell, 1993) and the Clinical Analysis Questionnaire (CAQ; Krug, 1980). These questionnaires have shown adequate validity in officer selection (Hart, 1981) and have demonstrated good predictive power for police officer performance in a number of outcomes such as supervisor ratings (Fabricatore, Azen, Schoentgen, & Snibbe, 1978; Schuerger, Kochevar, & Reinwald, 1982) or academy achievement (Topp & Kardash, 1986; Waugh, 1996).

LEADR scores have shown good test-retest reliability, with an average of 0.83 in a short-term interval (up to 2 days) and one of 0.64 in a long-term interval (more than 4 years). The report has shown good criterion validity based on various job performance criteria, such as direct evaluation from superiors, commendations, tenure, and disciplinary sanctions (IPAT, 1987). LEADR scores have also yielded good discriminating power: Scores on several dimensions have been found to differ significantly in employed law enforcement officers and non-law enforcement populations. The LEADR performance index yields higher values in law enforcement officers than in the general population and higher even than in military personnel—a result that indicates that the report is sufficiently sensitive to discriminate performance in tasks specifically related to the duties of law enforcement officers.

Although there are exceptions, law enforcement performance prediction studies have mainly focused on predicting performance at the police academy. Studies that concentrate on personality measures are no exception. The reason is obvious: Assessment processes in this context are designed to screen out potentially unsuccessful applicants and thus maximize the efficiency of the training courses. Arguably though, the main goal of performance prediction is to predict not training outcomes but actual job performance after graduation. In the case of police officers, this is a matter of public concern.

To achieve this goal, the potential relationships and predictive power of personality scores should not be isolated from the context where the skills are developed, that is, from academy training. In this study our main aim was to examine the predictive validity of the whole training process as well as personal dispositions such as personality and motivation on actual professional performance. To do so, we gathered information from several sources (probations, reprimands, and supervisor ratings) to construct a latent job performance proxy variable. A secondary aim was to provide data on police officer selection in a non-English-speaking context.

METHOD

PARTICIPANTS

The sample comprised 2,010 police candidates enrolled in the Catalan Institute for Public Safety (ISPC), assessed and trained during the years 1999 through 2001, who successfully completed the full training course. Cohort sizes were as follows: 1999, $n = 670$; 2000, $n = 694$; 2001, $n = 644$. These cohorts of candidates comprise the full population during this period. Average age of the full sample was 25.4 years. All were European; women accounted for 18.6% ($n = 374$) of candidates and men for 81.4% ($n = 1,636$). The minimum educational level required to be accepted in the ISPC is high school or equivalent (46.5% of the sample; $n = 935$), and the highest educational level attained in this population was college (53.5% of the sample; $n = 1,075$). There were no significant differences between cohort groups concerning gender, age, or educational level. The assessment process is compulsory for all candidates, and so in the context of the selection and training process, they received no compensation for their participation in the study.

THE ASSESSMENT PROCESS

All candidates completed the 16PF-5 (Cattell et al., 1993) and CAQ (Krug, 1980). Scores on these two measures may contribute to LEADR dimensions either positively (i.e., higher scores contribute to higher LEADR dimension scores) or negatively (i.e., lower scores contribute to higher LEADR dimension scores). The dimensions are:

1. Emotional Adjustment: Overall adjustment and coping style as suggested by the pathology-oriented scales of the 16PF and the CAQ.
2. Integrity/Control: Organization, planning, and ability to adapt to group standards.
3. Intellectual Efficiency: Judgment ability in situations requiring problem solving.
4. Interpersonal Relations: Information about the style of relationship with others.

Table 1 displays 16PF and CAQ descriptive statistics as well as their relation with the LEADR scores. Two additional LEADR subscales were used: accident/error proneness (AC) and leadership (LE). LEADR scales were reported in a sten metric (standardized scores with a mean of 5.5 and standard deviation of 2) allowing comparisons with observational values.

The CAQ (Krug, 1980) was used because of empirical evidence demonstrating that clinical information is relevant for behavioral adjustment to law enforcement tasks (Hargrave & McKinley, 1989; Inwald, Knatz, & Shusman, 1982; Shusman, Inwald, &

TABLE 1: Descriptive Statistics for 16-PF and CAQ Subscales and Their Weight in LEADR Report

	Scale	Average Score	SD	Weight in LEADR	
				Positive	Negative
16-PF	A	16.60	4.19	IR	P
	B	10.57	2.33	EA, IE, P	
	C	18.65	3.36	EA	
	E	14.32	3.05	EA	
	F	12.74	3.19	EA, IR	P
	G	18.03	3.83	EA, IC, P	
	H	17.72	4.33	EA, IE	
	I	7.27	4.06		EA, P
	L	2.52	3.59		EA, IE
	M	2.83	2.56		EA
	N	7.30	4.52	EA, P	
	O	3.33	3.35		EA, IE, P
	Q1	2.19	4.36		
	Q2	0.89	2.55	IE, IR	IR
	Q3	16.47	3.84	EA, IC	
	Q4	1.85	3.19		EA, IE
CAQ	D1	0.45	0.82		EA
	D2	0.64	1.38		EA
	D3	10.13	2.30		EA
	D4	5.30	2.14		EA
	D5	0.19	2.20		EA
	D6	2.49	2.60		EA
	D7	1.26	1.62		EA
	PA	1.48	2.54		EA
	PP	12.73	2.32		EA
	SC	2.26	1.38		EA
	AS	6.29	1.74		EA
	PS	4.08	2.02		EA

Note. 16PF = 16 Personality Factor Questionnaire; CAQ = Clinical Analysis Questionnaire; LEADR = Law Enforcement Assessment and Development Report; LEADR scale names: EA = Emotional Adjustment; IC = Integrity/Control; IE = Intellectual Efficiency; IR = Interpersonal Relations; P = Performance; 16 PF scale names: A = Warmth; B = Reasoning; C = Emotional Stability; E = Dominance; F = Liveliness; G = Rule Consciousness; H = Social Boldness; I = Sensitivity; L = Vigilance; M = Abstractedness; N = Privateness; O = Apprehensions; Q1 = Openness to Change; Q2 = Self-Reliance; Q3 = Perfectionism; Q4 = Tension; CAQ scale names: D1 = Hypochondriasis; D2 = Suicidal Depression; D3 = Agitated Depression; D4 = Anxious Depression; D5 = Low Energy Depression; D6 = Guilt and Resentment; D7 = Boredom and Withdrawal; PA = Paranoia; SC = Schizophrenia; AS = Psychasthenia; PS = Psychological Inadequacy.

Landa, 1984). This questionnaire has been used in police selection with good results (Bishop et al., 2001; Lorr & Strack, 1994), and its application is mandatory for computing the Emotional Adjustment scale of the LEADR. Additional predictive questionnaire information was based on the CAQ second-order factors Depression, Psychoticism, Sociopathy, Neuroticism, and Anxiety.

Behavioral data are a key component of the ISPC assessment process. Behavior information was compiled using an assessment center format, which was conducted during academy training for the purposes of both training and evaluation. Assessment centers have been found to be useful for selection purposes in law enforcement contexts (Coulton & Feild, 1995; see also Aamodt 2004b). In accordance with suggestions by several authors (Arthur, Day, McNelly, & Edens, 2003; Hough & Oswald, 2000; Lievens, 1998), a board of experts designed an observational coding protocol along with an operative definition of

the variables to be measured. The following aspects of police training were appraised during practical exercises using this protocol:

1. Training Attitude: Suitability of attitude and behavior in different training situations.
2. Job Efficacy: Level of achievement during the completion of a training task.
3. Motivation for Police Tasks: Attitude toward improvement in a police task.
4. Responsibility: Disposition toward the acceptance of duties and obligations.
5. Practical Judgment: Accuracy when judging situations and facts.
6. Initiative and Autonomy: Ability to conduct a task without help or support.
7. Adaptation to Norms: Compliance with the ISPC rules and orders.
8. Integration in the Team: Ability to work collaboratively as part of the team.
9. Social Skills: Ability to establish effective communication and relationships.
10. Tolerance/Flexibility: Acceptance of other people's attitudes, opinions, and behaviors.

All these aspects were evaluated at the ISPC by three trainers during three different simulation sessions, each one with specific settings and objectives. Final qualifications were reported on a 10-point scale, ranging from *unacceptable* to *excellent*. The qualification was based on an observational record protocol developed for the assessment of each setting. This protocol showed good indexes of agreement for all 10 subscales (average kappa = 0.92).

At the end of academy training, candidates underwent a final evaluation to summarize their performance during training. Two aspects were evaluated. First, candidates were graded according to their behavioral performance during an exercise. The ISPC used the same observational record protocol as the previous behavioral exercises for the assessment. Second, academic performance was evaluated averaging exam scores for the different courses: human rights, statutory law, traffic law, criminal law, administrative law, public safety system, sociology, geography, first aid, criminal investigation, police intervention, report briefing, mediation in conflicts, and theory contents of self-defense training. All candidates had to achieve a grade of 5 or more points on both the observational and the academic exam to graduate from the ISPC. Candidates with qualifications below 5 points on either exercise were excluded from subsequent training.

Another performance predictor was based on the results of an interview conducted at the end of training. During this interview, psychologists evaluated the trainee's ability to cope with stressful situations. Besides an extensive written report for each interview, the final appraisal was ranked in three ordinal categories of predicted adjustment: low (3.1% of the total sample), medium (21.1%), and high (75.7%).

PERFORMANCE CRITERIA

After academy training, successful candidates spent 1 year of real-life practice on at least one professional assignment. After this year of training, graduates received a supervisor rating summarizing their performance during the year, consisting of an overall score evaluating several aspects of performance. This score (PGA-01) was based on a questionnaire covering the aspects assessed during the academy assessment center and had a theoretical range between 0 and 162. In addition, supervisors were asked a general checking question (QPGA-01): "In which group do you consider the candidate's performance should be classified: (1) bottom 25%; (2) between 25% and 75%; (3) top 25%?"

Supervisors were asked the same checking question 6 years after graduation (QPGA-07). Note that fully trained, experienced police officers were being evaluated at this point in time.

We also recorded the number of disciplinary sanctions, disciplinary proceedings, and probations received by candidates at the end of the 6-year period as well as the number of commendations for excellent performance and medals and individual or team commendations registered in the service record. Table 2 displays zero-order correlations between the different predictor measures and outcomes.

DATA ANALYSIS

We fitted a structural equation model (SEM) to compute the amount of variance explained by LEADR and observational records using raw scores. Regarding the performance criteria, the number of commendations and sanctions and the scores on the PGA questionnaires were treated as quantitative data variables. Checking questions QPGA-01 and QPGA-07 were treated as ordinal data.

The main structural equation modeling was conducted using MPLUS 4.1 (Muthén & Muthén, 2006) with a diagonally weighted least squares estimator (DWLS; Christoffersson, 1977), using mean and variance corrections (Satorra & Bentler, 1994) for the fit indexes, so that the analysis provided exact *p* values. This estimator is denoted as WLSMV in MPLUS. Although demographic variables were initially considered, their inclusion in the model did not yield substantial differences in the fit or prediction power, so they were removed for the sake of parsimony.

RESULTS

DESCRIPTIVE RESULTS

All LEADR and observational variables were converted to a common sten metric. Table 3 displays means, medians, and minimum and maximum scores for graduates on the variables studied.

The average number of commendations and sanctions was very low (see Table 3), and the frequency distribution of these outcomes was highly skewed, meaning that the continuous measure was not of great use for analytical purposes. However, the award of at least one commendation (variable commendation) or at least one sanction (variable sanction) was not an uncommon event over the 6-year period. We therefore transformed these variables into dichotomous variables providing information on the presence/absence of these performance criteria. As a result, 15.6% of the sample had at least one commendation and 45.2% had at least one sanction or reprimand during the 6-year period.

Interview reports were also treated as dichotomous variables. The categories low performance and medium performance were collapsed because the frequency of the first category was just 3.1%. Interview reports were thus analyzed dichotomously, using a joint low category (24.3%) and the original high category (75.7%).

PGA-01 and PGA-07 continuous scores were rejected for modeling purposes due to their restricted range and low variance. Concerning the dependent variables, the QPGA-01

TABLE 2: Correlations Between 16PF, CAQ, and Performance Variables

Table with 30 columns (A-SAN) and 30 rows (A-SAN). The table contains correlation coefficients between 16PF, CAQ, and performance variables. Significant correlations are marked in boldface.

Note N = 2,010. Significant correlations (|r| < .05) are marked in boldface. 16PF = 16 Personality Factor Questionnaire; CAQ = Clinical Analysis Questionnaire; 16 PF scale names: A = Warmth; B = Reasoning; C = Emotional Stability; E = Dominance; F = Liveliness; G = Rule Consciousness; H = Social Boldness; I = Sensitivity; L = Vigilance; M = Abstractedness; N = Phriendness; O = Openness to Change; Q2 = Self-Reliance; Q3 = Perfectionism; Q4 = Tension; CAO scale names: D1 = Hypochondriasis; D2 = Suicidal Depression; D4 = Anxious Depression; D5 = Low Energy Depression; D6 = Guilt and Resentment; D7 = Bottom and Withdrawal; PA = Pianist; SC = Schizotypy; AS = Psychastenia; PS = Psychological Inadequacy; AQ = academic qualification; BO = behavioral qualification; PGA.01 = overall score evaluating several aspects of performance; PGA.07 = overall score evaluating several aspects of performance 6 years later; SAN = presence of sanctions; MEN = presence of commendations; OPGA.01 = general checking question; OPGA.07 = general checking question 6 years later.

TABLE 3: Descriptive Statistics for Self-Reported, Behavioral, and Outcome Measures in the Study

		<i>Mean</i>	<i>Median</i>	<i>Standard Deviation</i>	<i>Minimum</i>	<i>Maximum</i>
LEADR	Emotional Adjustment	5.46	6.00	1.91	1.00	10.00
	Integrity/Control	5.48	6.00	1.97	1.00	10.00
	Intellectual Efficiency	5.43	6.00	2.01	1.00	10.00
	Interpersonal Relations	5.48	6.00	1.98	1.00	10.00
	Performance	5.48	5.00	1.96	1.00	10.00
	Accidents	5.44	5.00	1.99	1.00	10.00
CAQ	Leadership	5.47	5.00	1.95	1.00	10.00
	Neuroticism	5.49	5.00	1.98	1.00	10.00
	Anxiety	5.46	5.00	1.99	1.00	10.00
Observational	Sociopathy	5.49	5.00	2.00	1.00	10.00
	Training Attitude	5.45	5.50	0.59	3.00	8.00
	Job Efficacy	5.30	5.00	0.63	2.50	7.50
	Motivation Police Task	5.36	5.00	0.58	3.00	8.00
	Responsibility	5.41	5.50	0.63	2.50	8.00
	Practical Judgement	5.30	5.00	0.64	2.00	7.50
	Initiative	5.24	5.00	0.59	2.00	8.50
	Adaptation to Norms	5.48	5.50	0.61	2.50	8.00
	Integration into Group	5.46	5.50	0.58	2.50	8.00
	Social Skills	5.28	5.00	0.62	2.00	8.00
Exam Qualification	Tolerance/Flexibility	5.32	5.00	0.56	3.00	8.00
	Academic	7.23	7.27	0.65	2.50	8.95
Criteria	Behavioral	5.74	5.75	0.57	2.87	7.98
	Number of sanctions	0.13	0.00	0.53	0.00	3.00
	Number of commendations	0.27	0.00	0.62	0.00	5.00
	PGA-01	122.45	123.00	6.76	94.00	153.00
	PGA-07	122.88	123.00	9.22	64.00	157.00

Note. LEADR = Law Enforcement Assessment and Development Report; CAQ = Clinical Analysis Questionnaire; PGA-01 = overall score evaluating several aspects of performance; PGA-07 = overall score evaluating several aspects of performance 6 years later.

checking question obtained 23.2% responses for the first option, 48.4% for the second, and 28.4% for the third. Using the QPGA-07, 23.0% of candidates were assigned to the first option, 50.2% to the second, and 26.8% to the third. Commendations and reprimands were infrequent. Due to the low occurrence and low standard deviation of the number of commendations and sanctions, they were not used as quantitative variables for analytical purposes, but as dichotomous indicators of latent performance. Commendations and sanctions, along with the QPGA-01 and QPGA-07, served as categorical indicators for a law enforcement performance latent variable generated by a measurement model inserted in the full performance model. The factor model for latent performance showed acceptable fit indexes: root mean square error of approximation (RMSEA) = 0.048, Comparative Fit Index (CFI) = 0.94, Tucker-Lewis Index (TLI) = 0.95.

SEM MODEL FOR POLICE PERFORMANCE

An important guideline when constructing the model was the sequence of the different measures. The assessment schedule was organized in three different, consecutive instances: (a) Phase I, comprising a questionnaire and observational assessment of candidates during instruction; (b) Phase II, consisting of the exams for the final grade and graduation at the

end of ISPC training; and (c) Phase III, postgraduation assessment once the candidates had completed training and were already engaged on a specific assignment. Thus, the final model respects the stages of the assessment process, guaranteeing that the model is really modeling performance temporally and making predictions in the same chronological sense.

Initially, a direct effects model from Phase I questionnaire data to law enforcement performance yielded very poor results, accounting for just 3% of the variance in law enforcement performance. The only significant predictors of this questionnaire model were Emotional Adjustment ($\beta = .17, p = .03$), Accident ($0.21, p = .02$), and CAQ-Anxiety ($\beta = .18, p = .02$). A direct effects model from Phase I observational data to law enforcement performance showed that the indicators did not result in significant predictions of latent performance (using $\alpha = .05$).

A second direct effects model was used with Phase II academic and behavioral qualifications as predictors of latent performance. This model accounted for 27.3% of performance and the paths showed that academic and behavioral qualifications were positively related with performance ($\beta_{\text{academic}} = .35, p < .01$; $\beta_{\text{behavioral}} = .28, p < .01$).

The final model gathered all the information in a sequential process. Questionnaire information and observational records from Phase I were added to the Phase II direct effects model. The final performance model is displayed in Figure 1, along with the standardized estimated parameters. The criteria used to assess overall model fit were (a) the RMSEA, which must be below 0.08 to show acceptable fit (Browne & Cudeck, 1993), and (b) the Comparative Fit Index and the Tucker-Lewis Index, which compare the covariation of the theoretical model to that of a null model (a model where all correlations or covariances are zero), with values above 0.90 indicating excellent fit to the data. In the final model, solid arrows represent significant paths ($p < .05$), and dashed arrows represent nonsignificant paths. This model has 231 degrees of freedom and shows excellent approximate fit (RMSEA = 0.051). Incremental fit indexes also yield excellent values for the model (CFI = 0.96, TLI = 0.95).

The full model showed outstanding predictive power for law enforcement performance, explaining 60% of the variance of this variable. Table 4 displays indirect and specific effects of predictors on individual criteria. Observational indicators had the greatest indirect effect on Law Enforcement Performance, with values ranging from .40 to .51. Initiative was the sole observational indicator that did not yield substantial effects for prediction purposes, but it must be noted that this was also the only indicator without a significant factor loading on the Observational factor. Questionnaire data showed lower indirect effects, although LEADR's Integrity/Control, Accidents, and Intellectual Efficiency demonstrated moderately high indirect effects, yielding values over .10. Interestingly, clinically oriented second-order factors Anxiety and Sociopathy from the CAQ yielded indirect effects of 0.15 and 0.13, respectively.

As expected, specific effects of predictors on performance indicators were smaller. Specific effects on QPGA-01 and QPGA-07 were similar, with slightly higher values for the first assessment (QPGA-01). Specific effects on Commendations were very close to zero, as explained by the low factor loading of this indicator on the latent performance variable ($\lambda = 0.11$). Predictably, specific effects on the indicator Sanctions were negative, although they were greater in magnitude than for any other indicator in light of their factor loading on the latent performance variable ($\lambda = -.39$).

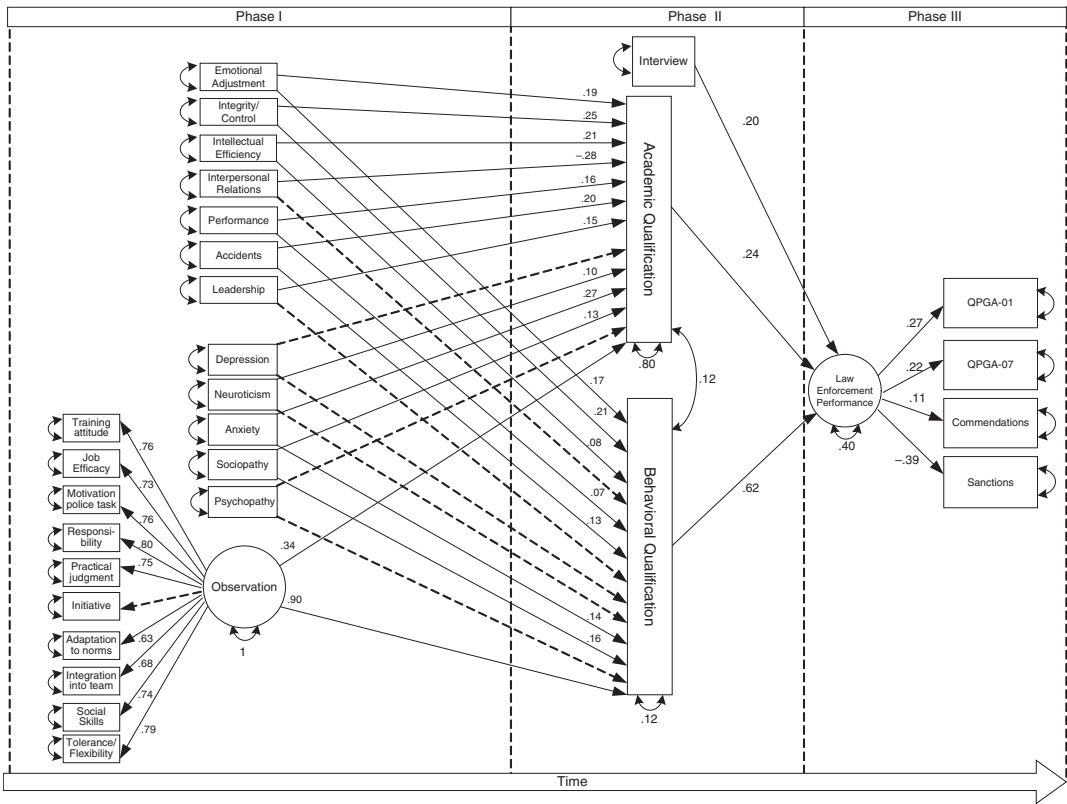


Figure 1: Final Structural Model for Law Enforcement Performance Prediction
 Note. Values display standardized parameters. Dashed arrows represent nonsignificant paths. QPGA-01 = general checking question; QPGA-07 = general checking question 6 years later.

DISCUSSION

Many ancillary studies of police performance have centered on predicting performance during training, based on personality and cognitive measures. In contrast, the present study focused on the long-term professional performance of successful graduates from the Catalan Institute of Public Safety, using a number of indicators in a three-stage assessment protocol. During Phase I candidates were assessed using the CAQ (Krug, 1980) and the 16PF-5 (Cattell et al., 1993) to compute LEADR report scores (IPAT, 1987). The LEADR is a set of composite scores devised by experts to serve as indices of performance in law enforcement environments. We also used a behavioral observation protocol specifically developed to appraise attitudes, dispositions, and interpersonal skills during training. Using this information, we developed a model for job performance with a measure indicated by two different sources: (a) reports from superiors and supervisors and (b) the presence of disciplinary action and/or commendations in the service record. The model included academy qualifications as a mediator of questionnaire measures.

TABLE 4: Indirect and Specific Effects on Law Enforcement Performance and Specific Effects on Criteria Indicators

		<i>Indirect Effect on Law Enforcement Performance</i>	<i>Specific Effect on QPGA-01</i>	<i>Specific Effect on QPGA-07</i>	<i>Specific Effect on Commendations</i>	<i>Specific Effect on Sanctions</i>
LEADR	Emotional Adjustment	.06	.02	.01	.01	-.02
	Integrity/Control	.19	.05	.04	.02	-.07
	Intellectual Efficiency	.10	.03	.02	.01	-.04
	Interpersonal Relations	.07	.02	.01	.01	-.03
	Performance	.08	.02	.02	.01	-.03
	Accidents	.13	.03	.03	.01	-.05
CAQ	Leadership	-.04	-.01	-.01	.00	.01
	Neuroticism	.02	.01	.01	.00	-.01
	Anxiety	.15	.04	.03	.02	-.06
Observational	Sociopathy	.13	.04	.03	.01	-.05
	Training Attitude	.49	.13	.11	.05	-.19
	Job Efficacy	.47	.13	.10	.05	-.18
	Motivation Police Task	.49	.13	.11	.05	-.19
	Responsibility	.51	.14	.11	.06	-.20
	Practical Judgment	.48	.13	.11	.05	-.19
	Initiative	.00	.00	.00	.00	.00
	Adaptation to Norms	.40	.11	.09	.04	-.16
	Integration into Group	.43	.12	.10	.05	-.17
	Social Skills	.47	.13	.10	.05	-.18
	Tolerance/Flexibility	.51	.14	.11	.06	-.20
	Observation (latent)	.64	.17	.14	.07	-.25

Note. QPGA-01 = general checking question; QPGA-07 = general checking question 6 years later; LEADR = Law Enforcement Assessment and Development Report; CAQ = Clinical Analysis Questionnaire.

Interestingly, the Phase I predictors accounted for 18.9% of the variance in academic qualification and 79.5% of the variance in behavioral qualification (these two variables being the Phase II predictors). These values are consistent with previous empirical research (Detrick et al., 2004). Phase I predictors showed good short-term power to evaluate performance at the academy, even when only candidates who graduated were considered. Nevertheless, a number of direct effects models using questionnaires and observational data as predictors did not find that these sources of information allowed good predictions of performance *after* academy training. For their part, Phase II predictors comprising only academy qualifications explained 27.3% of the latent performance variable. As was to be expected, academy training had a crucial effect on candidates' subsequent performance. This result validates the success of the training program.

In order to make valid predictions, we gathered all possible information in a sequential SEM model that included Phase I and Phase II predictors. In this model, personality and observational measures were mediated by academy training performance to predict actual job performance. This model thus integrated the effects of candidates' dispositions and motivation into training results. The key concept was that the model not only tried to predict training results but also put these results together to predict actual professional performance.

Officer performance information was gathered from sources that spanned the 6 years after graduation. Both the data gathering process and the fitted model were longitudinal: In the model, each phase contributed to the prediction of all the subsequent phases.

Consequently, questionnaire variables allowed a certain amount of prediction from Phase I to Phase II, particularly for academic qualification. Even though a direct effects model from Phase I to Phase III did not yield substantial prediction power, it greatly improved prediction accuracy when mediated by final training performance inside the academy. These results suggest that actual job performance is indeed influenced by psychological variables but that these influences pass through training. Hence, the models presented here suggest that training is a mediator of personal dispositions. In practice, this result would imply that psychological assessment in law enforcement personnel selection should be conducted with an eye on the academy training program. In our study, direct influences of personality on actual performance were weak; their true effect was captured much better when mediated by training. Thus, the capacity of police officers seems to depend on how their psychological dispositions are modulated by formal training. Whether this training effect is restricted to this specific job environment or is a generalized phenomenon of mediation is not discussed here, although it seems an interesting topic for further research.

All questionnaire measures were found to be significant predictors of academic qualification, but not all of them were related to behavioral qualification. Of all LEADR measures, Integrity/Control was the best positive predictor of academic qualification ($\beta = .25$), whereas Interpersonal Relations was the highest negative predictor ($\beta = -.28$). Interpersonal Relations is a composite related to Extraversion. Although Extraversion is often positively correlated with performance in early educational periods (Entwistle, 1972), it has been found to be negatively related with performance after secondary stages of education (O'Connor & Paunonen, 2006). It is worth noting that Intellectual Efficiency and Performance, two LEADR subscales associated with cognitive ability and execution, had similar, positive weights in academic prediction (around .20). Integrity/control was also the LEADR subscale most associated with behavioral qualification ($\beta = .21$). Other subscales associated with this qualification were Emotional Adjustment ($\beta = .17$) and CAQ Sociopathy ($\beta = .17$). On the whole, these values give the impression that successful police officers tend to be less anxious and more controlled than their less successful counterparts. This result is in agreement with those of other studies, which suggest a profile of low emotional vulnerability and high conscientiousness as a predictor of performance in law enforcement samples (Detrick et al., 2004; Piedmont & Weinstein, 1998). This personality pattern may be associated with better coping styles in law enforcement job environments (Bishop et al., 2001).

Major efforts have been devoted to producing ecologically valid measures at the ISPC, especially for observational protocols. The development of practical exercises and the assessment of variables that are sensitive to task adjustment have received significant human and economic investment. These efforts have clearly paid dividends; in our study, observational measures of trainees' behavior turned out to be excellent predictors of Phase II and Phase III outcomes. The latent observational factor is a very good predictor of both academic qualification ($\beta = .34$) and behavioral qualification ($\beta = .90$). This factor seems to reflect a fairly unspecific dimension of adjusted functioning during training. As the factor loadings indicated, the major effects come from Responsibility, a scale designed to measure dispositions related to Conscientiousness and Openness to Experience factors of the NEO-PI (Costa & McCrae, 1992). As we mentioned earlier, Integrity/Control was one of the questionnaire's chief predictors of academy performance and, indirectly, of professional performance after graduation. This result is in line with previous research supporting this

dimension as a core predictor of general job performance (Barrick & Mount, 1991; Piedmont & Weinstein, 1998; Salgado, 1997; Tett et al., 1991) and law enforcement performance (Bishop et al., 2001; Black, 2000; Cortina, Doherty, Schmitt, Kaufman, & Smith, 1992; Detrick et al., 2004). Another interesting result in the observational subscales concerns Tolerance/Flexibility, which was also related to Openness to Experience. On the other hand, Training Attitude and Motivation for Police Tasks were motivational measures of individual commitment to training and to an agent's various duties. In any case, all these observational measures were of great value for predicting both qualifications (except for initiative, which had a nonsignificant factor loading, $\lambda = 0.02$, on the Observational factor).

The focus of this study was to fit a longitudinal model for predicting law officer performance several years after training. Previous research has provided sound evidence that personality dimensions are predictors of performance in police training (Aamodt, 2004b; Barrick & Mount, 1991; Black, 2000; Detrick et al., 2004; Tett et al., 1991; Varela et al., 2004). Here, we propose a model that goes beyond mere association studies and proposes a sequential, causal path for postacademy performance that includes training results and personal dispositions as predictors.

Nevertheless, the fact that the study does not take into account cognitive functioning measures is an important shortcoming when predicting long-term performance. Cognitive abilities have proved to be essential measures in performance prediction in law enforcement selection (Aamodt, 2004b) and in other contexts (e.g., Colom, Escorial, Shih, & Privado, 2006). Cognitive abilities must be included in any consideration of the capacity of a police cadet because a police officer's everyday tasks require reading comprehension, attention to detail, verbal aptitude, and so on. Integrating these measures into the model would surely provide a great deal of information on future performance. As previous research has noted, cognitive abilities would be expected to correlate with job performance in general (Barrick & Mount, 1991).

Another weakness of this study is that all models are devoted to detecting potential high-performance trainees. Current research in law enforcement personnel selection concentrates heavily on screening out potentially unsuccessful or problematic applicants (for extensive coverage of this topic, see Aamodt, 2004a, 2004b). Incorporating unsuccessful applicants and trainees into the model would provide much fuller information by comparing their measures with those of high-performance and low-performance graduates.

Despite these limitations, our findings suggest that it is possible to predict high-performance police officers accurately at very early stages of their training. Personality is a major predictor, but its effect is mediated by training. Our results have clear implications for training design and personnel selection.

REFERENCES

- Aamodt, M. G. (2004a). *Law enforcement selection: Research summaries*. Washington, DC: Police Executive Research Forum.
- Aamodt, M. G. (2004b). *Research in law enforcement selection*. Boca Raton, FL: Broken Walker Publishing.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

- Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18*, 359-372.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9-30.
- Bishop, G. D., Tong, E. M. W., Diong, W. M., Enkelmann, H. C., Why, Y. P., Khader, M., et al. (2001). The relationship between coping and personality among police officers in Singapore. *Journal of Research in Personality, 35*, 353-374.
- Black, J. (2000). Personality testing and police selection: Utility of the "Big Five." *New Zealand Journal of Psychology, 35*, 353-374.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Cattell, R. B., Cattell, A. K., & Cattell, H. E. P. (1993). *Sixteen Personality Factor Questionnaire* (5th ed.). Champaign, IL: Institute for Personality and Ability Testing, Inc.
- Christofferson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika, 40*, 433-438.
- Colom, R., Escorial, S., Shih, P. C., & Privado, J. (2006). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42*, 1503-1514.
- Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*, 339-346.
- Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "Big Five" personality factors in the IPI and the MMPI: Predictors of police performance. *Personnel Psychology, 45*, 119-140.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costello, R. M., Schoenfeld, L. S., & Kobos, J. C. (1982). Police applicant screening: An analogue study. *Journal of Clinical Psychology, 38*, 216-221.
- Coulton, G. F., & Feild, H. S. (1995). Using assessment-centers in selecting entry-level police officers: Extravagance or justified expense? *Public Personnel Management, 24*, 223-254.
- Detrick, P., Chibnall, J. T., & Luebbert, M. C. (2004). The revised NEO personality inventory as a predictor of police academy performance. *Criminal Justice and Behavior, 31*, 676-694.
- Drees, S. A., Ones, D. S., Cullen, M. J., Spilberg, S. W., & Viswesvaran, C. (2003, April). *Personality assessment in police officer screening: Mandates and practices*. Paper presented at the 18th Annual Meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.
- Entwistle, N. (1972). Personality and academic performance. *British Journal of Educational Psychology, 42*, 137-151.
- Fabricatore, J., Azen, S., Schoentgen, S., & Snibbe, H. (1978). Predicting performance of police officers using the Sixteen Personality Factor Questionnaire. *American Journal of Community Psychology, 6*, 63-69.
- Funder, D. (2008). Persons, situations and person-situation interactions. In O. John, R. Robins, & L. Pervin (Eds.), *Handbook of personality* (3rd ed., pp. 568-582). New York: Guilford.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135-164.
- Hargrave, G. E., & McKinley, J. C. (1989). Use of the California Psychological Inventory in law enforcement officer selection. *Personnel Psychology, 53*, 267-277.
- Hart, R. (1981). *The use of the clinical analysis questionnaire in the selection of police officers: A validation study*. Unpublished doctoral dissertation, Florida State University.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10*, 249-254.
- Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—Remembering the past. *International Journal of Selection and Assessment, 51*, 631-664.
- Inwald, R. E., Knatz, H., & Shusman, E. (1982). *The Inwald Personality Inventory Manual*. New York: Hilson Research.
- IPAT. (1987). *The Law Enforcement Assessment and Development Report (LEADR). Report manual*. Champaign, IL: The Institute for Personality and Ability Testing.
- IPAT & Staff. (1998). The Law Enforcement Assessment and Development Report (LEADR) *Security Journal, 10*, 171-181.
- Krug, S. E. (1980). *Clinical Analysis Questionnaire manual*. Champaign, IL: Institute for Personality and Counseling.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6*, 141-152.
- Lorr, M., & Strack, S. (1994). Personality profiles of police candidates. *Journal of Clinical Psychology, 50*, 200-207.
- Mischel, W. (1968). *Personality and assessment*. New York: John Wiley.
- Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resource management. *Research in Personnel and Human Resources Management, 13*, 153-200.
- Muthén, B., & Muthén, B. (2006). MPLUS (Version 4.1) [Computer software]. Los Angeles: Author.
- O'Connor, M. C., & Paunonen, S. V. (2006). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*, 971-990.

- Ozer, D. J., & Benet-Martínez, V. (2005). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401-421.
- Piedmont, R. L., & Weinstein, H. P. (1998). Predicting supervisor ratings of job performance using the NEO Personality Inventory. *Journal of Psychology, 128*, 255-265.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82*, 30-43.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications to developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Schuerger, J. M., Kochevar, K. F., & Reinwald, J. E. (1982). Male and female corrections officers: Personality and performance. *Psychological Reports, 51*, 223-228.
- Sellbom, M., Fischler, G. L., & Ben-Porath, Y. S. (2007). Identifying MMPI-2 predictors of police integrity and misconduct. *Criminal Justice and Behavior, 34*, 985-1004.
- Shusman, E. J., Inwald, R. E., & Landa, B. (1984). Correction officer job performance as predicted by the IPI and MMPI: A validation and cross-validation study. *Criminal Justice and Behavior, 11*, 309-329.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.
- Topp, B. W., & Kardash, C. A. (1986). Personality, achievement and attrition: Validation in a multiple-jurisdiction police academy. *Journal of Police Science & Administration, 14*, 234-241.
- Varela, J. G., Boccaccini, M. T., Scogin, F., Stump, J., & Caputo, A. (2004). Personality testing in law enforcement employment settings. *Criminal Justice and Behavior, 31*, 649-675.
- Waugh, L. (1996). *Police officer recruit selection: Predictors of academy performance*. Queensland, Australia: Queensland Police Academy.

Carlos G. Forero is assistant professor of psychology at the University of Barcelona. His research interests focus on studying the efficiency of estimation methods in categorical factor analysis and on the application of these methods in risk assessment and the prediction of consequential outcomes. He has received the Society of Multivariate Experimental Psychology dissertation project award.

David Gallardo-Pujol is assistant professor of psychology at the University of Barcelona. His research interests fall into two main areas: the conceptualization and measurement of impulsive aggression and the experimental study of gene-environment interactions in the development of personality traits. He also works as an applied psychologist, being focused on the prediction of behavior using personality and intelligence data in forensic and educational contexts.

Alberto Maydeu-Olivares is associate professor of psychology at the University of Barcelona. His research interests focus on developing new quantitative methods for psychological applications. He has received the American Psychological Association dissertation award (Division 5), the Catalan Young Investigator Award, and the Society of Multivariate Experimental Psychology Cattell's award. He is currently Section Editor of *Psychometrika*.

Antonio Andrés-Pueyo is currently a full professor of individual differences at the University of Barcelona (UB). His research covers a wide range of interests, from psychophysiology of personality, to the relationship between smoking and personality, or more recently, the prediction of violence risk. He leads the Violence Studies Research Group at UB, which recently has adapted some violence risk management tools (SARA, SVR, HCR-20) to Spanish and Catalan.