The reliability of the behavioral-personnel assessment device (B-PAD) in selecting police recruits

William G. Doerner and Terry M. Nowell
School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA

Keywords *Recruitment, Selection, Testing, Police, Assessment*

Abstract *A recent entry into the police selection tools market has been the behavioral-personnel assessment device (B-PAD). Applicants view videotaped scenarios and are instructed to respond as if they were the officer handling the situation. Participant responses are preserved on tape for subsequent review by panels of three raters. Given the instrument's limited reliability checks to date and literature that questions the reliability of the oral board process, the present study examines whether B-PAD ratings are influenced by the demographic backgrounds of raters and ratees. The data were based on 113 subjects exposed to eight scenarios graded by panels of three judges. The results indicate that B-PAD scores are independent of rater race and sex, as well as ratee race and sex. While the B-PAD appears to be reliable and free from bias, the study cautions that a need for an independent validity check still remains.*

The budgetary outlay associated with the recruitment, selection, and training of new officers represents a formidable burden for many local law enforcement agencies. Adverse publicity, costly legal battles, and federal consent decrees have forced administrators to modify entrance procedures to ensure a more demographically representative candidate pool (Martin, 1991). Despite this reorientation, hiring practices continue to focus on "weeding out" unacceptable applicants rather than "screening in" desirable neophytes. While this whittling process eventually pares the field down to a more manageable size, a fundamental flaw underlies this approach. That is, survivors are not necessarily the "cream of the crop". They merely represent applicants whom the agency has found no reason to reject.

Despite concerted efforts to eliminate unacceptable job seekers from the applicant pool, an inefficient hiring process becomes further marred by premature employee turnover. Early departures mean that agencies are not able to recoup their recruitment costs, let alone realize any gains on their initial investments. For instance, some states now require training academy graduates to pass an independently administered examination to gain certification (Doerner, 1997a). Causalities at this point mean that agencies must absorb the price of refilling these slots and limp along beneath their authorized strength until suitable replacements are on board.

The authors wish to thank the editor and the anonymous reviewers for providing a series of very thoughtful and insightful comments that guided the revision process.

Policing: An International Journal of Police Strategies & Management, Vol. 22 No. 3, 1999, pp. 343-352.

Post-academy training programs, whether by design or as a result of lingering resentment and antagonism (Doerner and Patterson, 1992; Fagan, 1985; Felkenes and Schroedel, 1993), are another drop-out point. Field training officer programs typically anticipate a 35 per cent attrition rate, although losses of up to 50 per cent are not uncommon (Doerner *et al.*, 1989; Felkenes *et al.*, 1993). Turnover can exasperate diligent EEOC recruitment efforts accomplished at the front end and undermine any benefits associated with retaining a pool of diverse seasoned officers (Doerner, 1995). At the same time, federal grants aimed at mobilizing 100,000 additional officers on the streets under the guise of community policing programs amount to a deliberate infiltration of "short-timers" into the rank-and-file (Buerger, 1996; Doerner, 1997b). As one can see, all these obstacles drive up the costs of inefficient personnel selection procedures and can hamper the attainment of hiring goals.

Agencies have made concerted efforts to combat this problem and streamline hiring procedures. Most departments rely upon polygraphing, physical wellness standards, educational requirements, medical reviews, and background checks to screen out patently unqualified applicants. In addition, they have incorporated psychological testing, personality typologies, cognitive examinations, interest inventories, oral board interviews, assessment centers, and situational tests to help evaluate candidate quality (Ash *et al.*, 1990; Burbeck and Furnham, 1985; Gaines and Falkenberg, 1998; Hogue *et al.*, 1994; Sanders *et al.*, 1995). Unfortunately, these techniques are time-consuming, costly, and, most importantly, do not enjoy the same degree of reliability or validity. As a result, agencies are left in a continuously ongoing search to incorporate more defensible devices into their personnel decision-making process.

One recent entry into the police selection tools market has been the behavioral-personnel assessment device (B-PAD). According to the B-PAD developers (http://www.bpad.com), more than 200 law enforcement agencies have processed over 30,000 applicants to date with this instrument. Evidence of the B-PAD's growing popularity can be gleaned from advertisements which grace the classified pages of one trade magazine and tout the benefits of pretest coaching materials (Anonymous, 1999).

Despite growing popularity and increased agency adoptions, it is not clear whether this product possesses any distinct advantages over other existing protocols. Only a limited number of reliability and validity checks have been undertaken. Furthermore, the instrument's developers, as opposed to independent researchers, have performed all the reliability and validity assessments to date. This state of affairs suggests that the time is ripe for this procedure to receive additional third-party scrutiny.

The behavioral-personnel assessment (B-PAD) device

The B-PAD was designed to assess an applicant's interpersonal skills and judgment (Corey *et al.*, 1995). The system consists of a variety of videotaped situations similar to what an actual officer could encounter while on the job.

For example, candidates might face an unruly child who is acting defiantly towards a parent, an impatient motorist who belittles the officer during a traffic stop, a distraught elderly lady who complains about being the repeated target of neighborhood teenage vandals, a quarreling husband and wife, a group of loitering and belligerent youths, a disoriented female who is possibly the victim of a sexual assault, and so forth. The applicant is instructed to assume that the actors in the scenes are real people and to respond to them as if he or she were dealing with the actual situation. After watching the tape, the applicant has a 45-second interval in which to make inquiries to the parties, solicit information, issue commands, and render advice about available options or impending actions. At the end of this period, the screen fades out and a new scenario is presented. This process is repeated until the candidate has reacted to a total of eight different scenes. The entire administration takes about 30 minutes to complete.

The vignettes and the participants' responses are preserved on tape for subsequent review. This practice circumvents most of the logistical problems associated with assembling a face-to-face oral interview board (Doerner, 1997a). Three trained raters review the videotaped performance and grade the applicant's responses in terms of task orientation, interpersonal skills, and overall effectiveness with each scenario. *Task orientation* reflects an assessment of the applicant's demonstrated problem-solving ability. *Interpersonal skills* conveys whether the candidate related appropriately to the actor(s) in the scene. Finally, *overall effectiveness* is a blend of these two dimensions. Scores range from 1 to 4, where 4 signifies the highest level of competency.

The instrument's developers report that the B-PAD is both a reliable and valid assessment device. In one unpublished study, 50 California law enforcement applicants took the B-PAD. Here it was determined that the B-PAD task orientation scale was reliable because it demonstrated a significant correlation with an independent measure of problem-solving ability ® = 0.39, $p < 0.01$). An unrefereed second study (Young, 1992) reported that deputy sheriff applicants who flunked a relevant entrance exam pertaining to communications skills also registered significantly lower B-PAD interpersonal skills scores. A third unpublished project (Rand, 1987) aimed at establishing concurrent validity. It compared overall effectiveness scores of 30 working police officers with supervisory ratings and found a significant correlation ® = 0.72, $p < 0.01$). Other unpublished analyses have determined that the B-PAD does not have an adverse impact on women and minorities, is not affected by a candidate's prior law enforcement experience nor by previous exposure to the testing device, correlates significantly with academy instructor ratings of recruits, and that overall effectiveness scores correlate significantly with promotional candidate responses to another independent measure (Corey *et al.*, 1995). In sum, these materials form the developers' basis for suggesting that the B-PAD is a reliable and valid assessment device for selecting entry-level police officers.

While considerable time and energy have gone into the construction and development of the B-PAD, these efforts ignore a parallel body of literature. A number of studies examine the oral interview board process at various junctures in the police world (police academy admissions, job entry, promotions) and consistently demonstrate that extraneous factors compromise candidate ratings (Doerner, 1997a; Falkenberg *et al.*, 1990; Gaines and Kappeler, 1992; Gaines and Lewis, 1982; Landy, 1976; Reynolds, 1979). Since the B-PAD is basically an extension of the oral interview board process in a slightly different technological guise, it is hard to envision how this instrument escapes the ensnaring perils uncovered in the police literature.

When these two lines of research are juxtaposed with each other, an apparent contradiction is exposed. On the one hand, it appears that the B-PAD developers have gone to great lengths to create a technically sound instrument. On the other hand, it seems that social processes routinely invade the oral board process and contaminate rater outcomes. One way to resolve this rift is for a third party to conduct an independent empirical analysis. Whether the strengths associated with the B-PAD can compensate for the flaws that typically envelop the oral board process is an empirical question that awaits resolution. As a result, the purpose of this research is to test for any discriminatory effects present within the B-PAD by examining the influence that race and gender of the applicants and raters exert upon B-PAD scores.

Study site

The study site is the Tallahassee (Florida) Police Department (TPD) which is located in the state capital. The city is home for state government, as well as two major universities and one large community college. The population is over 130,000 inhabitants and the city encompasses an area of more than 70 square miles. The department itself has an authorized strength of 320 sworn positions. TPD underwent its initial accreditation in 1986 and has been reaccredited three more times since then.

TPD fell under a federal consent decree during the early 1980s. The agreement stipulated that the agency needed to concentrate more efforts on the recruitment and hiring of minorities and women. That legal settlement heralded the introduction of many personnel changes, including the establishment of a field training officer (FTO) program and a revamped officer selection process.

Study group and variables

The initial study group consisted of 118 subjects who submitted employment applications for sworn positions to TPD between June 1997 and June 1998, survived a cursory review of their credentials to ensure compliance with the advertised requirements, and were invited to take the B-PAD test. The B-PAD process contains seven different versions of scenarios. Inspection of the database showed that 113 applicants were exposed to either of two forms. The remaining five candidates were distributed among three other variations. As a

result, these five cases were eliminated from further analysis. Perusal of the data revealed that two applicants were Hispanic and four applicants fell into the "other" racial category. These files were eliminated from subsequent analysis to allow the study to concentrate on the racial contrast between blacks and whites.

Preliminary analyses revealed that the independent variables exhibited similar distributions in both video forms. In the first form (n = 38), 29 per cent of the ratees were black and 76 per cent were male, compared to corresponding figures of 30 per cent and 77 per cent in Form B (n = 69). Rater characteristics showed that 15 per cent were black and 51 per cent were male in Form A (n = 114), while Form B had a 16 per cent and 52 per cent representation on race and sex, respectively (n = 207). Thus, the choice of the form had no effect upon the independent variable distribution.

Table I presents the correlation coefficients between the three rating scales (task orientation, interpersonal skills, and overall effectiveness) for each scenario contained within the two forms. Each candidate was graded by three raters, yielding a total of 114 ratings for the 38 subjects exposed to Form A and 207 ratings for the 69 subjects exposed to Form B. Inspection of the table reveals a high degree of congruity, as one might expect, between the overall effectiveness scale and the other two rating scales. As mentioned earlier in the section describing the B-PAD device, overall effectiveness is a blend of the scores earned on task orientation and interpersonal skills. Given the conceptual composition of the overall effectiveness scale and the high intercorrelations, it was decided to retain just the grades on the overall effectiveness scale for further presentation.

The purpose of the B-PAD is to "screen in" superior candidates rather than simply "weed out" inferior candidates. At the time these data were collected,

	Form A			Form B		
	Scale 1[a] Scale 2	Scale 1 Scale 3	Scale 2 Scale 3	Scale 1 Scale 2	Scale 1 Scale 3	Scale 2 Scale 3
Scene 1	0.67	0.88	0.80	0.68	0.89	0.81
Scene 2	0.77	0.87	0.89	0.59	0.85	0.75
Scene 3	0.66	0.85	0.80	0.69	0.87	0.82
Scene 4	0.72	0.89	0.83	0.80	0.94	0.87
Scene 5	0.76	0.88	0.89	0.72	0.92	0.82
Scene 6	0.85	0.93	0.90	0.82	0.91	0.91
Scene 7	0.80	0.88	0.90	0.77	0.89	0.89
Scene 8	0.80	0.90	0.90	0.76	0.92	0.85
Average r	0.75	0.89	0.86	0.73	0.90	0.84
Number of subjects		38			69	
Number of raters		114			207	

Notes: [a]Scale 1 refers to the ratings for task orientation, Scale 2 pertains to scores issued on interpersonal skills, and Scale 3 indexes grades for overall effectiveness.

Table I. Pearson correlation coefficients between scale ratings by form and scene

TPD had extended job offers to 18 candidates and had declined to hire 86 job seekers. Focusing on persons who had received a job offer and persons whom the agency had rejected revealed that the successful applicants had garnered significantly higher B-PAD scores than the rejected group. Furthermore, 22 per cent of the 912 ratings issued under Form A and 36 per cent of the 1,656 scores involving Form B landed in the highest rating category. As a result of this information, it was decided to construct the dependent variable so that it represented the allocation of superior scores (a value 4 on the rating scale) versus less than superior scores (values of 1, 2 or 3 on the rating scale). When this operation was carried out, the eight scenarios in Form A posted 18 per cent, 14 per cent, 14 per cent, 18 per cent, 30 per cent, 35 per cent, 19 per cent and 25 per cent of the dependent variable scores in the high category. The corresponding figures for Form B were 28 per cent, 23 per cent, 25 per cent, 22 per cent, 46 per cent, 47 per cent and 44 per cent.

Results

As mentioned earlier, the variables of interest include the race and sex of both the raters and the ratees. An intercorrelation matrix using Pearson r coefficients (not displayed here) was constructed. For Form A, five of the six coefficients landed in single digits while the largest relationship (ratee race and ratee sex = 0.19) attained significance at the 0.05 level. For Form B, none of the intercorrelations were significant and the largest of the six coefficients reached the 0.11 mark. Thus, the risks associated with exposure to multicollinearity are minimal.

The initial multivariate analysis indicated that which form of the B-PAD was administered emerged as a significant predictor. A peek at the dependent variable distributions displayed in the previous section underscores this point. These observations led the researchers to partition the vignettes. Scenarios were categorized according to their subject matter. Situations that called for a law enforcement response were separated from clips that were more service-oriented or portrayed officer conduct. For example, a law enforcement response was appropriate in scenes that depicted a mutually combative husband-and-wife domestic battery, a flag-down where a young hitchhiker reports a sexual molestation, a traffic stop involving a motorist who belittles the officer, and a group of men engaged in an argument which is about to become physical. Instances that called for a service response involved a parent unable to handle an unruly child, a death notification, a suicide-in-progress call initiated by a neighbor hearing shots fired, an elderly lady being badgered by neighborhood teenagers, and a mentally unstable individual causing a disturbance in a public building. Scenes that portrayed the discovery of a sergeant drinking while on duty, an officer applying more force than necessary during an arrest, an officer stealing an item while checking out an unsecured business, a sergeant who starts screaming while disciplining an officer for a minor policy violation, and a distraught officer in a post-shooting case fell into the officer behavior category.

Separating the scenes in this manner allows a more intensive way to comb the data for any possible intrusions associated with the demographic backgrounds of raters and ratees.

Table II summarizes the logistic regression solutions for the enforcement-related scenarios. None of the predictors attain significance in any of the scenarios. Furthermore, none of the models is significant and the amount of variation explained is minimal. Thus, rater and ratee characteristics have no bearing on applicant scores in the enforcement vignettes.

Table III presents the logistic regression solutions when the scenes depict a social service response. None of the predictors reaches significance in any of the scenarios. Only one model, Scenario 5, is significant. With that one exception, the amount of variation explained is minimal. Once again, demographic variables exert no impact on rater assessments of applicants.

Table IV presents the logistic regression solutions when the scenes deal with officer behavior. Only one predictor in the entire table is significant. Only one model, Scenario 3, is significant. With that one exception, the amount of variation explained is minimal. Thus, it would appear that ratings are independent of social attributes.

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
Variable	B s.e.	Odds Ratio	B s.e.	Odds Ratio	B s.e.	Odds Ratio	B s.e.	Odds Ratio
Rater race[a]	0.44	1.56	0.59	1.80	0.34	1.41	–0.85	0.43
	(0.52)		(0.43)		(0.39)		(1.10)	
Rater sex	–0.09	0.91	–0.20	0.82	–0.15	0.86	0.22	1.25
	(0.40)		(0.34)		(0.27)		(0.62)	
Ratee race	–0.30	0.74	–0.01	0.99	–0.14	0.87	–1.75	0.17
	(0.43)		(0.34)		(0.29)		(1.07)	
Ratee sex	0.60	1.82	0.60	1.83	0.56	1.75	–0.51	0.60
	(0.44)		(0.42)		(0.37)		(1.15)	
Rater race X	–0.05	0.95	–0.14	0.87	–0.08	0.93	2.54	12.63
Ratee race	(0.92)		(0.74)		(0.67)		(1.86)	
Rater sex X	–0.17	0.84	0.38	1.46	–0.18	0.84	0.52	1.69
Ratee sex	(0.75)		(0.60)		(0.54)		(1.45)	
Constant	–1.31*		–1.55*		–0.67*		–1.56*	
	(0.31)		(0.26)		(0.21)		(0.47)	
N	207		321		321		114	
–2 log likelihood	217.528		319.068		411.093		87.310	
Goodness of fit	205.455		321.449		320.901		111.181	
Cox and Snell R^2	0.020		0.030		0.017		0.044	
Nagelkerke R^2	0.031		0.047		0.023		0.080	
% Class. correctly	77.29		79.13		65.11		85.96	

Notes:
*Denotes significance at the 0.05 level of analysis.
[a]Variable coding: Race (0 = black, 1 = white), Sex (0 = female, 1 = male).

Table II.
Logistic regression coefficients representing estimated effects for raters issuing a superior score in enforcement-related scenarios

Variable	Scenario 1 B s.e.	Odds Ratio	Scenario 2 B s.e.	Odds Ratio	Scenario 3 B s.e.	Odds Ratio	Scenario 4 B s.e.	Odds Ratio	Scenario 5 B s.e.	Odds Ratio
Rater race[a]	–0.67	0.51	0.67	1.95	–0.30	0.74	1.07	2.90	–1.16	0.31
	(0.60)		(0.50)		(0.48)		(0.71)		(1.10)	
Rater sex	–0.21	0.81	–0.38	0.68	–0.16	0.86	–0.21	0.81	–0.74	0.48
	(0.36)		(0.33)		(0.31)		(0.54)		(0.57)	
Ratee race	–0.11	0.89	–0.51	0.60	0.24	1.28	–1.83	0.16	–0.56	0.57
	(0.37)		(0.35)		(0.34)		(1.07)		(0.62)	
Ratee sex	0.24	1.28	0.57	1.76	0.09	1.09	–0.95	0.39	–0.10	0.91
	(0.48)		(0.46)		(0.45)		(1.14)		(0.76)	
Rater race X	0.50	1.65	–0.12	0.88	0.01	1.01	–7.02	0.00	1.19	3.30
Ratee race	(0.95)		(0.83)		(0.80)		(39.05)	(1.65)		
Rater sex X	–0.01	0.99	–0.13	0.87	–0.52	0.59	–6.67	0.00	0.19	1.21
Ratee sex	(0.73)		(0.69)		(0.68)		(25.13)	(1.17)		
Constant	–0.80*		–0.01		0.13		–0.98*		–0.84*	
	(0.28)		(0.25)		(0.25)		(0.41)		(0.39)	
N	207		207		207		114		114	
–2 log likelihood	243.421		275.423		283.832		90.310		107.920	
Goodness of fit	207.348		206.709		207.033		91.282		111.187	
Cox and Snell R^2	0.010		0.050		0.014		0.151		0.033	
Nagelkerke R^2	0.015		0.067		0.019		0.245		0.053	
% Class. correctly	71.98		58.45		53.14		81.58		80.70	

Table III.
Logistic regression coefficients representing estimated effects for raters issuing a superior score in service-related scenarios

Notes:
*Denotes significance at the 0.05 level of analysis.
[a]Variable coding: Race (0 = black, 1 = white), Sex (0 = female, 1 = male).

Discussion and conclusion

The results of the present inquiry support the assertion that the B-PAD device is a reliable instrument when the focus of the investigation is on the demographic composition of both raters and ratees. In other words, the scores that applicants receive are independent of their racial and sexual characteristics and also are unrelated to the racial and sexual backgrounds of the raters. This conclusion remains constant for all the scenarios, regardless of whether they are oriented towards a law enforcement, social service, or officer conduct theme. Thus, the B-PAD ratings studied here do not appear to be contaminated by any unwarranted discriminatory biases.

While the B-PAD appears to be a reliable mechanism, the question of its validity remains an open and elusive concern. Unfortunately, the current database does not permit an independent validity assessment at this time. As mentioned earlier, the data collection period for the present study extended from June 1997 to June 1998. During that period, TPD rejected 86 job seekers and extended job offers to 18 candidates. These hires were staggered in three waves which coincided with the start of the training academy (October 1997, March 1998, and July 1998). The six-month academy curriculum, along with the five months consumed by the post-academy Field Training Officer Program,

Variable	Scenario 1 B s.e.	Odds Ratio	Scenario 2 B s.e.	Odds Ratio	Scenario 3 B s.e.	Odds Ratio	Scenario 4 B s.e.	Odds Ratio	Scenario 5 B s.e.	Odds Ratio
Rater race[a]	0.49	1.63	–0.22	0.81	1.42*	4.15	0.30	1.35	–0.35	0.71
	(0.52)		(0.48)		(0.70)		(0.66)		(0.84)	
Rater sex	–0.48	0.62	–0.06	0.94	–0.41	0.67	0.10	1.11	–0.17	0.85
	(0.40)		(0.32)		(0.47)		(0.46)		(0.52)	
Ratee race	–0.25	0.78	–0.60	0.55	–0.08	0.92	–0.84	0.43	0.03	1.03
	(0.43)		(0.35)		(0.56)		(0.53)		(0.53)	
Ratee sex	0.21	1.24	0.63	1.89	–2.03	0.13	0.07	1.08	0.27	1.31
	(0.50)		(0.46)		(1.10)		(0.70)		(0.71)	
Rater race X	–0.10	0.90	0.78	2.18	–0.15	0.86	0.29	1.33	0.69	2.00
Ratee race	(0.92)		(0.81)		(1.22)		(1.19)		(1.28)	
Rater sex X	0.14	1.15	–0.06	0.94	0.97	2.64	0.39	1.48	–0.09	1.09
Ratee sex	(0.78)		(0.68)		(1.39)		(0.95)		(0.99)	
Constant	–1.11*		–0.09		–0.57		–0.57		–1.12*	
	(0.00)		(0.25)		(0.36)		(0.36)		(0.39)	
N	207		207		114		114		114	
–2 log likelihood	214.980		279.066		125.846		143.931		126.114	
Goodness of fit	206.781		206.853		118.562		113.604		113.903	
Cox and Snell R^2	0.021		0.032		0.121		0.033		0.009	
Nagelkerke R^2	0.032		0.043		0.171		0.045		0.13	
% Class. correctly	77.78		56.04		74.56		65.79		75.44	

Notes:
*Denotes significance at the 0.05 level of analysis.
[a]Variable coding: Race (0 = black, 1 = white), Sex (0 = female, 1 = male).

Table IV.
Logistic regression coefficients representing estimated effects for raters issuing a superior score in officer-related scenarios

means that it takes just about a year to put a rookie in uniform and out on the streets. Given this time frame and the small sample size, one can understand why a validity check utilizing these subjects is not feasible at this time.

The current researchers are not aware of any third-party published studies linking B-PAD success with subsequent training appraisals or actual on-the-job performance. While the relative youthfulness of this instrument may be a primary reason for this gap, the fact remains that critical employment decisions are being reached without the proper foundation. For the time being, it would behove agencies to shield themselves against potential litigation by demanding such evidence rather than adopting this protocol blindly.

References

Anonymous (1999), "Free book for police applicants", *Police: The Law Enforcement Magazine*, Vol. 23 No. 4, p. 58.

Ash, P., Slora, K.B. and Britton, C.F. (1990), "Police agency officer selection practices", *Journal of Police Science and Administration*, Vol. 17 No. 4, pp. 258-69.

Buerger, M. (1996), "Police education and the crime bill", *Police Forum*, Vol. 6 No. 1, pp. 1-16.

Burbeck, E. and Furnham, A. (1985), "Police officer selection: a critical review of the literature", *Journal of Police Science and Administration*, Vol. 13 No. 1, pp. 58-69.

Corey, D.M., MacAlpine, D.S., Rand, D.C., Rand, R. and Wolf, G.D. (1995), *B-PAD Technical Reports*, The B-PAD Group, Inc., Napa, CA.

Doerner, W.G. (1995), "Officer retention patterns: an affirmative concern for police agencies?", *American Journal of Police*, Vol. 14 No. 3/4, pp. 197-210.

Doerner, W.G. (1997a), "The utility of the oral interview board in selecting police academy admissions", *Policing: An International Journal of Police Strategies & Management*, Vol. 20 No. 4, pp. 777-85.

Doerner, W.G. (1997b), "Recruitment and retention", in Dantzker, M.L. (Ed.), *Contemporary Policing: Personnel, Issues, & Trends*, Butterworth-Heinemann, Boston, MA, pp. 53-73.

Doerner, W.G. and Patterson, E.B. (1992), "The influence of race and gender upon rookie evaluations of their training officers", *American Journal of Police*, Vol. 11 No. 2, pp. 23-36.

Doerner, W.G., Speir, J.C. and Wright, B.S. (1989), "An analysis of rater-ratee race and sex influences upon field training officer program evaluations", *Journal of Criminal justice*, Vol. 17 No. 2, pp. 103-14.

Fagan, M.M. (1985), "How police officers perceive their field training officer", *Journal of PoliceScience and Administration*, Vol. 15 No. 1, pp. 138-152.

Falkenberg, S., Gaines, L.K. and Cox, T.C. (1990), "The oral interview board: what does it measure?", *Journal of Police Science and Administration*, Vol. 17 No. 1, pp. 32-9.

Felkenes, G.T. and Schroedel, J.R. (1993), "A case study of minority women in policing", *Women and Criminal Justice*, Vol. 4 No. 2, pp. 65-89.

Felkenes, G.T., Peretz, P. and Schroedel, J.R. (1993), "An analysis of the mandatory hiring of females: the Los Angeles police department experience", *Women and Criminal Justice*, Vol. 4 No. 2, pp. 31-63.

Gaines, L.K. and Falkenberg, S. (1998), "An evaluation of the written selection test: effectiveness and alternatives", *Journal of Criminal Justice*, Vol. 26 No. 3, pp. 175-83.

Gaines, L.K. and Kappeler, V.E. (1992), "Selection and testing", in Cordner, G.W. and Hale, D.C. (Eds.), *What Works in Policing? Operations and Administration Examined*, Anderson Publishing, Cincinnati, OH, pp. 107-23.

Gaines, L.K. and Lewis. B.R. (1982), "Reliability and validity of the oral interview board in police promotions: a research note", *Journal of Criminal Justice*, Vol. 10 No. 5, pp. 403-19.

Hogue, M.C., Black, T. and Sigler, R.T. (1994), "The differential use of screening techniques inthe recruitment of police officers", *American Journal of Police*, Vol. 13 No. 2, pp. 113-24.

Landy, F.J. (1976), "The validity of the interview in police officer selection", *Journal of Applied Psychology*, Vol. 61 No. 2, pp. 193-8.

Martin, S.E. (1991), "The effectiveness of affirmative action programs: the case of women in policing", *Justice Quarterly*, Vol. 8 No. 4, pp. 489-504.

Rand, R.R. (1987), "Behavioral police assessment device: the development and validation of an interactive, pre-employment, job-related, video psychological test", *Dissertation Abstracts International 48(3 -A)*, pp. 610-611, University Micro films No. 8713282.

Reynolds, A.H. (1979), "The reliability of a scored oral interview for police officers", *Public Personnel Management*, Vol. 8 No. 5, pp. 324-8.

Sanders, B., Hughes, T. and Langworthy, R. (1995), "Police officer recruitment and selection: a survey of major police departments in the U.S.", *Police Forum*, Vol. 5 No. 4, pp. 1-4.

Young, T. (1992), "Use of the behavioral police assessment devices in the selection of law enforcement officers", doctoral dissertation, the Professional School of Psychology, University Microfilms No. LD02540.