

**SIFTING THE GOLD FROM THE PEBBLES.: USING SITUATIONAL  
INTERVIEWS TO  
SELECT CORRECTIONAL OFFICERS FOR NEW GENERATION JAILS**

Mary K. Stohr-Gillmore  
Political Science Department  
Washington State University

Michael W. Stohr-Gillmore  
Political Science Department  
Washington State University

Nannette Kistler  
Criminal Justice Program  
(Graduated MA 1988)  
Washington State University

Presented at the annual meeting of the  
Academy of Criminal Justice Sciences in  
Washington D.C., April 1989

No citation without the permission of the authors.

\* The authors wish to thank Dr. Nicholas P. Lovrich, Director of The Division of Governmental Studies and Services at Washington State University, for his assistance.

Acknowledgement: The research reported here was supported by a grant from the National Institute of Corrections, U.S. Department of Justice. The NIC bears no responsibility, however, for any omissions or errors which may be reported herein.

## **Introduction**

A growing body of research literature suggests that the arrival over a decade ago of podular/direct supervision New Generation” jails on the correctional horizon constituted a watershed in the management and supervision of inmate populations. The innovative podular architectural design, coupled with direct supervision of inmates by staff, has resulted in the development of facilities that are cost efficient, staff “job enriched” and safer and more humane for inmates (Nelson, 1986, 1987; Zupan et al., 1986; Zupan and Stohr-Gillmore, 1988). Previous work done by Menke et al. (1986) and Manning et al. (1988) would suggest that the success of direct supervision is contingent on several controllable organizational level factors. Among these factors is the development of selection processes for correctional officers which are capable of discriminating between those job applicants who are talented and able (the gold) and those who are more mediocre and less qualified (the pebbles). This paper examines whether a critical incident, behaviorally-based situational interview process, helps New Generation jail managers identify the requisite talents in prospective officers.

### ***New Generation Jails***

Since its inception, the American jail has been plagued by innumerable difficulties. The American jail as a public institution has demonstrated a poor track record regarding the ability to incarcerate citizens suspected and/or convicted of crimes without overcrowd-

ing, without noise, without rampant violence, without abuse of power by staff, without the development of a collateral inmate subculture, and without mind-numbing boredom. This systematic failure to incarcerate in a safe and humane fashion reached a point where the federal and state courts intervened in the internal affairs of executive branch law enforcement and correctional agency affairs—a step seldom taken by the American bench (Reid, 1976; Clear and Cole, 1986; Stohr-Gillmore, 1987).

Given this backdrop, the evolution of the podular/direct supervision “New Generation” jail represents a milestone in institutional corrections. The innovative concepts of podular architecture and direct supervision management of inmates are the basic components of these jails, making them far different in operation from the traditional jail’s linear architecture and intermittent supervision of inmates. In fact, some proponents of the New Generation jail concept claim that the podular/direct supervision jail’s underlying philosophy may in time revolutionize incarceration. In place of the violent, degrading, crowded, boring, dilapidated and poorly supervised traditional jail, the podular/direct supervision jail gives hope of a humane, safe and secure domicile for inmates (Gettinger, 1984).

Through direct proactive supervision of inmates: the podular/direct supervision jail: in contrast to the traditional facility, seeks as a main objective to insure the safety of staff and inmates (Gettinger: 1984). The threat of and actual wielding of physical coercion: while still available, is rarely resorted to as compliance is gained through persuasion and remuneration. Inmates in podular/direct supervision jails have more to “lose” by failing to comply in terms of the benefits of autonomy within their living unit, recreation, easy access to telephones, visitors and T.V. Therefore, one would expect that inmates

in podular/direct supervision jails would tend, in their own self-interest, to incite fewer disturbances.<sup>1</sup>

The New Generation jail concept does, then, signify a new role for the correctional officer. Instead of dealing with inmates in a removed and intermittent fashion, a correctional officer in the New Generation jail is in the same “pod” (secure dormitory-like area containing approximately forty inmates) with the inmates 24 hours a day. Rather than a job that is routine, fragmented and menial, the correctional officer’s job in the New Generation jail requires considerable supervisory, leadership and communication skills (Zupan, 1987). In fact, the job of the correctional officer in the New Generation jail might be termed “enriched” by the standards set by Hackman et al. (1981). That is, the correctional officer in the New Generation jail may find more satisfaction in his/her job than the correctional officer in a traditional jail because the New Generation jail correctional officer’s job is meaningful, provides direct responsibility for outcomes: and gives regular feedback about performance on the job (Zupan and Menke, 1987). Because of the new job responsibilities and skills required of correctional officers in New Generation jails, it is imperative that correctional managers have a selection process that screens effectively for such talented individuals.

### ***Selection Processes***

The Supreme Court ruled in *Griggs v. Duke Power Company (1971)* that “...an employer could not use a selection technique having an adverse impact on minorities unless that technique had been shown to measure job-related skills” (Thomas and Heisel,

---

1. Much of the information in the above on the New Generation Jails was taken from the unpublished thesis of Stohr-Gillmore (1987).

1983).<sup>2</sup> Because of the Griggs ruling, public managers are faced with a selection process dilemma: (1) continue with use of standardized but unvalidated tests and risk lawsuits if it can be shown that those tests have an adverse impact on minority groups; or, (2) resort to unstructured interviews that have no adverse impact, but which are also likely to represent poor selection practice for the identification of quality personnel. Students of selection processes make note that many public managers have chosen the latter course of action in the interest of meeting affirmative action goals. Instead of developing selection tests that can be validated, these employers have elected to take the “course of least resistance” and have returned to the interview format for employee selection to the detriment of employee quality (Daniel, 1986).

**The courts' and the Equal Employment Opportunity Commission's (EEOC) stringent requirements have reduced employment testing efforts in the United States without necessarily improving the position of women and minorities in the labor force. The argument that EEO standards have upgraded selection by stimulating increased validation may be true for a few organizations, but for others the strict requirements have had a chilling effect leading to the abandonment of old tests and an unwillingness to develop new ones. (Daniel, 1986: 1)**

As a consequence of the widespread return to the old unstructured interviews, some authors (e.g., Dyer, 1981) argue that the quality of the workforce will deteriorate primarily because interviews are not very good in discriminating between ‘those applicants who prove to be good and those who prove to be poor employees. Specifically, scholars are skeptical of the utility of interviews for predicting applicant success on most kinds of jobs (Wagner, 1949; Mayfield, 1964; Ghiselli, 1966; Latham and Saari, 1980; Silverman and Wexley, 1987), they question the utility of face-to-face interviews for gathering true impressions of personality characteristics (Ulrich and Trumbo, 1965), and they decry the

---

2. “Under the new guidelines (Equal Employment Opportunity Commission’s), a selection rate for any minority group which is less than 80 percent of the rate for the group with the highest rate will generally be regarded as evidence of adverse *impact*’ (Robertson, 1981).

generally “nebulous and intangible character” of most interview processes (Ghiselli, 1966: 389; Dyer, 1981).

As should be clear from the foregoing, the *Griggs* ruling clearly placed public personnel managers in a fix. If they continued to use unvalidated tests they risked discrimination charges, but if they returned to subjective interviewing techniques they risked hiring persons who would be incapable of performing well on the job. It is in this context of an unenviable choice between a rock and a hard place that the proponents of the “situational interview” process laud their technique as the solution to this personnel dilemma. According to Latham and Wexley (1982), a proper selection procedure is important both because it alleviates concerns about adverse impact (it can be validated) and because it assists organizations in screening for quality employees.

Before a proper situational interview can be constructed, however, the effective behaviors for a particular job must first be identified.

**A valid selection test cannot be developed until the organization agrees upon an acceptable definition (i.e., measure) of employee behavior. This is because the validity of a test is determined by measuring the performance of people on the test and measuring the performance of the same people on important aspects of the job. If there is a significant correlation between these two measures the selection procedure is valid (Latham and Wexley, 1982: 3).**

### ***The Critical Incident Technique***

The critical incident technique was developed by Flanagan (1954) and is utilized in the job analysis process by personnel specialists to identify the effective and ineffective behaviors associated with a given job (Flanagan, 1954; Latham and Wexley, 1982). In this technique those persons (job incumbents and supervisors) who currently perform or have recently observed the performance of the job in question are asked to give examples of **effective** and **ineffective** behaviors related to that job. Upon collection of these critical job

behaviors, researchers distribute the behaviors under logical dimensions-such as “Managing the living unit to assure a safe and humane environment”<sup>3</sup> in New Generation Jails (Latham and Wesley, 1982). In order to minimize sampling bias, a new representative group of job incumbents and supervisors are then contacted and asked to assign relative weights of importance to the dimensions specified, to categorize the effective and ineffective behaviors into the several dimensions, and to indicate the relative importance of each behavior under each job dimension (Latham and Wexley, 1982). From these weightings and categorizations the job analyst is able to construct a “situational interview” and a performance appraisal instrument which are intrinsically related to effective behavior on the job.

### ***Development of the Situational Interview Instrument***

The situational interview is devised as a job-related, structured and standardized procedure.’ That is, the same questions are asked of all applicants, the questions posed to applicants are job-related, and the applicants’ answers are benchmarked (Menke et al., 1986). This regimented process is followed in order to avoid the wide range of problems associated with unstructured interviews (e.g., rambling, digression, failure to cover all important material, lack of uniform treatment of applicants, etc.), to exclude legally prohibited discriminatory questions that are unrelated to job performance, to include “real life” decisions to be made on the job in the interview process, and to increase the likelihood of achieving a high degree of interobserver reliability between raters (Wagner,

---

3. This is an actual dimension developed for correctional officers under a grant from the National Institute of Corrections for Spokane County, Washington under the title, “Personnel Administration in New Generation Jails.”

1949; Mayfield, 1964; Latham and Saari, 1980; Menke et al., 1986; Silverman and Wexley, 1987; Maurer and Fay, 1988). The candidates are asked how they would react to several critical incidents; their answers are then compared to the most effective and least effective behaviors for dealing with those incidents identified by job occupants and supervisors. The situational interview technique is premised on the assumption that intentions are related to behavior (Silverman and Wexley, 1987) -that is, that what an applicant says s/he will do in the selection interview is “predictive” of actual subsequent job behavior.

### ***Development of the Situational Questions***

The situational questions and the benchmarked answers for each dimension were created through a three-stage group process developed by Latham et al. (1980: 422-427; Menke et al., 1986).<sup>4</sup> In the first stage of the process participants individually reviewed the dimensions and behaviors developed through the critical incident job analysis, and each one selected the three behaviors they personally felt best represented the dimension in question. The group then reviewed the individual selections and came to a consensus on the one behavior which best represented each dimension. Participants then described actual situations on the job where they had observed the critical behavior selected for each dimension. One situation was then selected by the group of participants which they felt best exemplified the critical behavior. The situation was then translated into a

---

4. The correctional facility personnel involved in this three-stage process included one lieutenant, sergeant and correctional officer at the Las Vegas Detention Center; one captain, lieutenant and sergeant at the Pima County Detention Center; and, one director and deputy chief at the Contra Costa Detention Center. “All of the participants had experience working as or supervising direct supervision correctional officers and most had experience interviewing correctional officer applicants. The inclusion of administrators, supervisors and line personnel ensured that a variety of perceptions and perspectives were represented in the interview questions and answers” (Menke et al., 1986: 36).



question like the one shown in the following section entitled “The Selection Process With the Situational Interview.” In order to construct the benchmarked answers, the group described responses to the situation that would be considered outstanding, tolerable and poor; consensus was achieved by the group of correctional personnel on each benchmarked answer.

## **Methodology**

The data for this study were collected from a New Generation jail in the North-western portion of the United States (hereafter referred to as the “Newjail”). The data is longitudinal in that the first group of applicants (collected in late **1985**) includes job applicants selected for the correctional officer job before the situational interview was included in the selection process; the second group (collected in early **1987**) includes job applicants selected with the assistance of the situational interview technique. The two groups of applicants included in this study were hired and represent the population of applicants hired during **1985, 1986** and 1987. In addition, each of these applicants were sent through the state correctional officers’ academy, completed a 15-day on-the-job training routine, and were evaluated on their job performance in the summer of 1988 and again in the fall of 1988.<sup>5</sup> The number of hires in the first group is 36, and the number of successful applicants in the second group is 33.

---

5. Both the summer and the fall evaluations were developed through the critical incident process. However, in the summer evaluation employees were rated only on the broad dimensions, whereas in the fall evaluation employees were rated on the fully developed correctional officer evaluation which includes the specific behaviors within the dimensions. In addition, in the fall evaluation, specialized forms (also developed through the critical incident process and in addition to the standard form developed for module officers) applicable to those officers assigned to booking, transport and the control room were used when appropriate.

## ***The Selection Process Without the Situational Interview***

The job screening process at the Newjail facility consisted of three general phases. In the first phase relevant background information was gathered on each applicant to screen out those individuals who did not meet the legal and ethical standards for work in a correctional facility. The second phase consisted of a written civil service exam and a physical fitness test. Both of these tests are intended to be job related and nondiscriminatory toward protected groups (Menke et al.: 1986). The third phase of the selection process is the interview phase. During this final stage a board of three interviewers makes the summary hiring recommendation to the civil service commission on the remaining applicants. The selection process for the two groups included in this study was very nearly identical, except, of course, for the interview phase.

In the first group interview condition, applicants were asked questions and rated by a board of three interviewers concerning some personal traits of the applicants thought important to the successful mastery and conduct of the position of correctional officer. The 11 rating categories of this phase included scores on the following: appearance and grooming, personality projection, temperament, poise, maturity, judgment, employment experience, police aptitude, military experience: career preparation and self confidence. In addition, a total rating was obtained by aggregating over all of the categories. Appended to each category were a few brief phrases describing the criteria for rating an applicant. For instance, the phrases “appears neat and well groomed” (high criteria score) and “spends little time on appearance” (low criteria score) were appended to the appearance and grooming category. The interviewer was instructed to consider these criteria as the extremes for a category, with a low criteria score warranting a rating of “one” and a high

criteria score a “ten.” The interviewers were instructed to rate the applicant between the one and ten extremes, using a five as the “average” individual standard.

### ***The Selection Process With the Situational Interview***

The second group interview condition required that the interviewers rate the applicant on six factors at several points in the interview-(1) after a review of the applicant folder, (2) after asking the applicant general questions, (3) during and (4) after asking the applicant five situational questions, (5) after the applicant has asked questions, and (6) after a discussion of the applicant among the interviewers and consensus is reached.<sup>6</sup> The

---

6. Training of the interviewers in how to use the situational questions was conducted in the month prior to the the actual hiring process. The training workshop consisted of several phases: (1) overview of the workshop objectives; (2) discussion of problems with interviewer bias; (3) discussion of effective interviewing guidelines; (4) instruction on Equal Employment Opportunity guidelines; (5) discussion of research on the use of situational questions; (6) A skill building exercise; and (7) a discussion and wrap-up. As part of the skill building exercise prospective interviewers practiced the use of the situational questions in a “mock interview format” on a recent correctional officer hire. After a practice reading of each of the original seven situational questions, each question was read aloud to the role play applicant. Responses were independently scored by the trainees and scores were shared and discussed to aid in the interpretation of each response against the behavioral anchors provided. Five of the seven questions were then chosen for use in the initial experimental use of this type of questioning. A follow-up training session was conducted six months later, after the interviewers had the chance to use the situational questions in “real life” selection situations, in order to identify any difficulties encountered when using the situational questions in the selection interview. During this session, the trainees raised three positive and two negative points. On the positive side: (1) the trainees felt that the situational question helped them spot “disciplinarian types” of applicants who lack effective interpersonal and “parenting skills” and who would be more likely to resort to one-on-one physical force; (2) the trainees felt that they were in better control of the interview content and were more likely to delay a positive recommendation to hire until the applicant had performed favorably on the situational questions; (3) the trainees also regarded the situational questions as explicit justification for rejecting a candidate. However, on the negative side: (1) the trainees felt the placement of the situational questions in the first section of the interview was awkward in that it didn’t allow for an “ice breaking period” (this was resolved by rearranging the interview format); (2) the trainees pointed out that some of the applicants might be receiving some coaching on the situational questions (this was resolved by expanding the pool of situational questions).

rating scale employed was a seven-point scale, with the “one” endpoint meaning unacceptable and the “seven” endpoint meaning outstanding. The six rating criteria used at five rating points were: communication skills, reasoning ability/judgment, interpersonal skills, relevant experience and education, interest in the position, and appearance. The questions employed at each point in the interview were structured, uniform and specific to jail work.

The rating scale for the situational questions was based on five points rather than the seven points utilized in the other scales. A five indicated that the applicant had given an outstanding response, a three indicated a mediocre or tolerable response, and a one indicated that the response was poor. The five situational questions used in each interview were related to a specific job situation that a correctional officer in a New Generation jail might encounter. Each question was “benchmarked” with possible answers that the correctional officer candidates might give; for some questions more than one answer might be benchmarked with the same number because the answers are different, but they have the same approximate value in the eyes of correctional personnel who participated in the creation of the questions. For example, the following question had two responses that were benchmarked with a value of one.

**Question #1: A facility rule states that inmates will be out of bed at 7:00 A.M. in the morning. In making your morning inspection of the module you notice that an inmate is still in bed. When you order him to get up he states that an officer on another shift has given him permission to stay in bed because he has a cold. What would you do?**

**5 = Verify the claim of the inmate by looking for documentation or contacting the officer.**

**3 = Call your supervisor and ask him/her whether its okay for the inmate to stay in bed or ask him/her what to do.**

**1 = Follow the rule of the facility without checking the inmate’s claim (make him get up).**

**1 = Don’t verify the claim and allow the inmate to stay in bed.**

### ***Measures of Behavior***

The two groups of hirees in this study participated in a 15-day on-the-job training course, an academy training course, and two performance evaluations. In the academy training the new hire was scored and evaluated on the following: competency in first aid, competency in interpersonal communication, report writing in the academy notebook, punctuality, interactions, attentiveness and appearance. In the on-the-job training the new hire was rated by one training officer on the following factors: personal appearance, personality, interest in work, quantity of work, quality of work, dependability, attitude, knowledge of job, and leadership qualities. An overall additive index score was calculated for both the academy training and for the on-the-job training.

The third measure of job performance was obtained by rating all participants on their job performance in June of 1988. Each participant was rated by a reviewing correctional supervisor on seven dimensions which were developed through the use of the critical incident technique. The reviewing supervisor was asked to rate each correctional officer on a scale of one to five on each dimension; one indicated a “poor” performance on the dimension, two indicates “below average”, three indicates “average”, four indicates “above average” and five indicated “outstanding” performance. For example, each correctional officer was evaluated under the “Resolving Inmate Problems and Conflicts” dimension with the following explanation appended: The extent to which the correctional officer provides guidance for the solution of inmate problems.<sup>7</sup>

---

7. The other six dimensions included in this evaluation were: Building Positive Rapport and Personal Credibility With Inmates, Maintaining Effective Administrative and Staff Relations, Managing the Living Unit to Assure a Safe and Humane Environment, Responding to Inmate Requests, Handling Inmate Discipline and Supervising in a Clear, Well-Organized and Attention-Getting Manner.

The fourth measure of job performance, like the third, was obtained by rating all participants on their job performance, but in the fall of 1988. Each participant was rated by a reviewing correctional supervisor on dimensions and their attendant behaviors developed through the use of the critical incident technique.<sup>8</sup> The reviewing supervisor was asked to rate each correctional officer with the appropriate form on each behavior by assigning a letter which represented the frequency with which a given behavior was observed by the supervisor. The letters A through E were utilized and defined thusly:

- A represents almost always or 95 to 100 percent of the time
- B represents frequently or 85 to 94 percent of the time
- C represents sometimes or 75 to 84 percent of the time
- D represents seldom or 65 to 74 percent of the time
- E represents almost never or 0 to 64 percent of the time

For example, the supervisor would choose one of the above designations when rating the officer on the following behavior: The correctional officer is courteous and polite when responding to public or other department inquiries. If the correctional officer is courteous and polite when responding to public or other department inquiries 95 to 100 percent of the time, the supervisor would circle A.

### ***Research Questions***

As established in the above, the correctional officer in the New Generation jail must

---

8. The number of dimensions and their attendant behaviors differed for those correctional officers evaluated with the “specialized” forms (e.g., booking, transport and control room officers). The “main” correctional officer form included seven dimensions and 56 behaviors, whereas the booking form included four dimensions and 36 behaviors, the control room form included five dimensions and 16 behaviors and the transport form included five dimensions and 32 behaviors.

possess skills commensurate with the requirements of an enriched and demanding job. Consequently, it is quite imperative that the selection process for such correctional facilities be tailored to select persons who possess the critical abilities demanded of the correctional officer job which requires the performance of continuous and prolonged direct supervision of forty or more inmates. Our main research, of course, is determining the degree to which the critical incident-based situational questions add to our ability to predict the scores of candidates in the academy, in the 15-day on-the-job evaluation, and in the performance rating earned after a year and more in service. We would expect to find evidence of a positive association between our dependent variables **On-the-job, Academy, Performance** and **Performance-2** (abbreviated as acad, perform and perform-2 in the tables) and our independent variables **Situational** and **After Situational** (abbreviated as sit and aftersit in the tables). More importantly, we would expect that the addition of the situational questions to the interview would result in an improvement in prediction of success in the academy, success on early job placement, and success on the job one year and more later. We expect to find that a comparison of overall board ratings for the two types of interviews (labeled **Total** for the first group and **Final** for the second group) would show that the second group produces a stronger positive relationship with the dependent performance variables.<sup>9</sup> Finally, research indicates that ratings based on discussion of the applicant among the panel of selection interviewers where consensus is reached yields higher validities (Silverman and Wexley, 1987). Therefore, we expected that the rating given after discussion and consensus is reached by the interviewers (in the

---

9. The independent variable Final represents cumulative ratings for the second group where the interviewer is asked to rate the applicant after the rest of the interview is completed. The first group cumulative rating (Total) is also arrived at by the interviewers after the interview has been completed.

second group only) would positively correlate with the dependent variables.

## **Discussion and Analysis**

### ***Findings***

Table 1 sets forth a comparison of personal and background characteristics for correctional officers in the group selected without the situational interview and the group selected with the situational interview. As illustrated by this table, the two groups are highly similar in composition, with only slight percentage differences on a few factors.



**Table 1**  
Distribution of Correctional Officer Characteristics

<i><b>Characteristic</b></i>	<i><b>First Group Without Questions</b></i>	<i><b>Second Group With Questions</b></i>
<b>NUMBER OF CASES:</b>		
N	36	33
<b>AGE:</b>		
Mean	33.9	34.0
Standard Deviation	6.8	10.4
<b>SEX:</b>		
Female	22.6%	27.6%
Male	77.4	72.4
<b>RACE:</b>		
White	90.3	93.1
Black	0.0	6.9
Hispanic	3.2	0.0
Other	6.5	0.0
<b>EDUCATION:</b>		
Less than high school	3.2	<b>0.0</b>
High school or GED	16.1	34.5
Some college, technical school or AA	58.1	58.6
Four year degree	19.4	6.9
Masters degree or above	3.2	0.0
<b>RELEVANT EXPERIENCE:</b>		
Military	16.1	24.1
Law enforcement	12.9	3.4
Security Officer	22.6	13.8
Correctional Experience	3.2	6.9
None	45.2	51.7

The relationship between the independent variables in the first group and the dependent variables is delineated in Table 2.<sup>10</sup> While only military experience appears strongly related (though not statistically significant) to success at the academy, temperament and poise are significantly correlated to the 15-day on-the-job evaluations and the appearance variable is significantly correlated with the performance variable.<sup>11</sup> The latter relationship is particularly troubling inasmuch as out of all of the independent variables included in the first group selection process, appearance is the **most vulnerable to legal challenge**. Despite the stated criteria for rating appearance (“appears neat and well groomed” or “spends little time on appearance”) use of the measure necessarily entails ratings based on attributes that individual applicants have relatively little control over—namely, their race, gender, age or physical impairment. The nebulous qualities of what constitutes “good looks” undoubtedly include those personal likes and dislikes that

---

10. The Pearson’s Correlation Coefficient ( $r$ ) is used here to determine the degree of association (with the exception of the last column). The reader should bear in mind when perusing these correlations that the range of applicants included in this study is necessarily restricted. The hiring process for the Newjail jail, as in most organizations, is a sifting process; those applicants who survive the first two phases of the process are graduated to the interviewer phase and, of course, only those who were hired were sent to the academy, trained on the job and given a performance evaluation. Therefore, since those persons who were hired ostensibly represent the “best” of those who applied for the job of correctional officer in Newjail, the correlations presented here may be correspondingly attenuated.

11. By the time the performance-2 evaluation was administered in the Fall of 1988, the numbers in each group had diminished through attrition to below 30. Given that the smaller the sample size the more likely that the results are to have occurred by chance and that a sample size smaller than 30 may be violating the assumption of a normal distribution (one of the assumptions which underlie the use of Pearson’s ( $r$ )) a supplemental measure of association, Kendall’s tau-c, is also included in this analysis. Kendall’s tau-c was chosen over other ordinal measures of association (e.g., Kendall’s tau-b and gamma) for a couple of reasons. First, tau-c is particularly suited over tau-b when “...dealing with variables having a different number of categories” (Weisberg, 1974: 1653). Secondly, the conventional wisdom is that gamma overstates the magnitude of a relationship, more so than tau-b or tau-c, when the  $N$  in the cells are low. Despite these justifications, however, the magnitude of all three in this analysis—tau-c, tau-b, and gamma—were remarkably similar.

make us all unique individuals, but which likely have little to do with doing a good job. Consequently, the need to develop a selection process that does not assign a major role to appearance becomes a matter of consequence.

Notably, none of the independent variables are strongly related to the performance-2 variable. In fact, five of those variables are negatively related. The tau-c measure of association indicates similar negative relationships between the “without-the-Situational-Interview Group” variables and the performance-2 variable.

---

**Table 2**  
Without-the-Situational-InterviewGroup<sup>†</sup>

	<i>Acad</i>	<i>On-the-job</i>	<i>Perform</i>	<i>Perform-2</i>	
N=36			N=31	N=26	tau-c
Appearance	-.136	.017	.455++	-.162	-.144
Personality	.071	.279	.267	.134	.106
Temperament	.002	.329++	.209	.010	-.006
Poise	-.004	.354++	.295	.171	.103
Maturity	.039	.245	.165	.041	.065
Judgement	-.227	-.194	.283	.044	.068
Experience	.217	.179	.076	-.068	-.108
Aptitude	.029	.195	.123	-.137	-.094
Military*	.369	.200	-.187	-.158	-.156
Preparation	-.038	.003	.084	-.118	-.063
Confidence	.131	.138	.294	.001	.077
Total	.114	.144	.148	-.044	-.037

---

<sup>†</sup>Pearson’s correlation (r) was used here with the exception of the last column.

++ Statistically significant at .05 level.

\*There were only 13 applicants with military experience.

The third table reports the strength of the linear relationships between the dependent variables used to assess performance with the situational interview group independent variables. These findings suggest that the individual situational question answers are far

less important than the composite after situational ratings<sup>12</sup> in predicting high academy performance, and they are also less important than the file, general, final or discussion ratings in predicting success on the 15-day initial job placement (On-the-job).<sup>13</sup> However, **the situational questions are more strongly related to success on the job a year or more later (Performance and Performance-2) than any other independent variable in this group.** Moreover, strong significant relationships between the Final and Discussion variables with the Performance-2 variables are also detected. Again, however, the Pearson's correlations between the second group variables and the Performance-2 variable must be viewed with caution due to a low sample number. The tau-c measure of association, though, further indicates the importance of the Discussion variable, and to a lesser degree, the General, Situational, After Situational and Final variables, as they relate to the Performance-2 dependent variable.

---

12. The after situational ratings are post-session perceptions gained by the interviewers as a result of hearing applicants respond to the entire set of five situational questions.

13. None of the independent variables are significantly correlated with the academy variable; the file and discussion variables are significantly correlated with the on-the-job variable, and only the situational variable is significantly correlated with the one-year-plus performance variable.

**Table 3**  
Situational Interview Group<sup>t</sup>

	<i>Acad</i>	<i>On- the-job</i>	<i>Perform</i>	<i>Perform-2</i>	
N=33			N=28	N=20	tau-c
File	.135	.312 <sup>++</sup>	-.129	.203	-.006
General	.155	.305	.125	.354	.204
Sit	.184	.004	.373 <sup>++</sup>	.482 <sup>++</sup>	.221
Aftersit	.249	.304	.227	.337	.228
Final	.156	.353	.189	.385 <sup>++</sup>	.260
Discussion	.197	.380 <sup>++</sup>	.154	.454 <sup>++</sup>	.288 <sup>++</sup>

<sup>t</sup>Pearson's correlation (r) was used here with the exception of the last column.

<sup>++</sup> Statistically significant at .05 level.

A comparison between the cumulative ratings' (Total and Final) predictive quality is presented in Table 4. As illustrated by this table, the Final rating was more highly correlated with all of the dependent variables than was the comparison Total ratings, indicating that the inclusion of the situational questions likely improves the overall predictability of the selection instrument for each of these performance measures employed.

**Table 4**

Summary Comparison of Interview Score Predictive Quality for With and Without Situational Questions

	<i>Acad</i>	<i>On-the-Job</i>	<i>Perform</i>	<i>Perform-2</i>
First Group				
Total	.114	.144	.148	-.044
Second Group				
Final	.156	.353	.189	.385 <sup>++</sup>

<sup>++</sup> Statistically significant at .05 level.

Tables 5 through 6 report the inter-rater agreement reliability coefficients (Pearson's correlation coefficient r) for the three raters in the first and second group interview condi-

tions. The inter-rater agreement reliability measure is assessed in order to determine the degree of consistency between raters achieved in evaluating all applicants. Low levels of agreement among raters is considered to constitute evidence that an assortment of interviewer biases might be operating (e.g., contrast error, “similar to me,” first impression, and halo) (Maurer and Fay, 1988). If this were the case, interview ratings likely would be reflective of something related to the interview structure or the interviewer as opposed to reflecting the ability of the applicant to do the job. When appraising an incumbent employee, a correlation of .60 is considered respectable; on this point Latham and Wesley argue that when “...the agreement is less than .60...it is likely that the appraisal is not measuring the employee’s performance, but rather the different attitudes and biases of the appraisers” (1982: 66). Given that the raters in this case were evaluating a job applicant based on a short interview rather than a history of employment spanning months or years, a correlation perhaps as low as .51 (correlation between Raters 2 and 3 in Table 5) should be considered acceptable. Of note here, the raters of the second group of applicants (the interviewer condition in which the situational questions were included) have higher agreement than the raters of the first group of applicants (the interviewer condition without the situational questions).

**Table 5**

Inter-rater Agreement Reliability Coefficients For the Without-the-Situational-Interview Group

	<i><b>Rater 1</b></i>	<i><b>Rater 2</b></i>	<i><b>Rater 3</b></i>
Rater 1		.680 <sup>++</sup>	.632 <sup>++</sup>
Rater 2	.680 <sup>++</sup>		.511 <sup>++</sup>
Rater 3	.632 <sup>++</sup>	.511 <sup>++</sup>	

<sup>++</sup> Statistically significant at .05 level.

**Table 6**

Inter-rater Agreement Reliability Coefficients For the Situational Interview Group

	<i><b>Rater 1</b></i>	<i><b>Rater 2</b></i>	<i><b>Rater 3</b></i>
Rater 1		.653 <sup>++</sup>	.711 <sup>++</sup>
Rater 2	.653 <sup>++</sup>		.548 <sup>++</sup>
Rater 3	.711 <sup>++</sup>	.548 <sup>++</sup>	

<sup>++</sup> Statistically significant at .05 level.

The inter-rater agreement reliability coefficients for select second group independent variables are displayed in Tables 7 through 10.<sup>14</sup> By comparison, the situational question raters (Table 7) and the discussion raters (Table 10) have the highest overall level of inter-rater agreement. One would expect high agreement scores between the discussion raters inasmuch as they are instructed to reach consensus on their final ratings. However, high agreement between the situational raters is a more unexpected finding, and may be indicative of the power of such questions to channel agreement between raters on applicant ability to perform the job.<sup>15</sup>

14. These variables were selected solely because they have the most relevance in terms of the impact of the situational questions on the interview process.

15. Maurer and Fay (1988: 339) made a similar discovery when they found that the situational interview is "...more effective than conventional structured interviews in producing agreement about job applicants among raters..."

**Table 7**

Inter-rater Agreement Reliability Coefficients For the Situational Questions

	<i><b>Rater 1</b></i>	<i><b>Rater 2</b></i>	<i><b>Rater 3</b></i>
Rater 1		.639 <sup>++</sup>	.645 <sup>++</sup>
Rater 2	.639 <sup>++</sup>		.416 <sup>++</sup>
Rater 3	.645 <sup>++</sup>	.416 <sup>++</sup>	

<sup>++</sup> Statistically significant at .05 level.**Table 8**

Inter-rater Reliability Coefficients For the After Situational Ratings

	<i><b>Rater 1</b></i>	<i><b>Rater 2</b></i>	<i><b>Rater 3</b></i>
Rater 1		.361 <sup>++</sup>	.643 <sup>++</sup>
Rater 2	.361 <sup>++</sup>		.261
Rater 3	.643 <sup>++</sup>	.261	

<sup>++</sup> Statistically significant at .05 level.**Table 9 .**

Inter-rater Agreement Reliability Coefficients For the Final Ratings

	<i><b>Rater 1</b></i>	<i><b>Rater 2</b></i>	<i><b>Rater 3</b></i>
Rater 1		.467 <sup>++</sup>	.665 <sup>++</sup>
Rater 2	.467 <sup>++</sup>		.321
Rater 3	.665 <sup>++</sup>	.321	

<sup>++</sup> Statistically significant at .05 level.



**Table 10**  
Inter-rater Agreement Reliability Coefficients For the Discussion Ratings

	<b>Rater 1</b>	<b>Rater 2</b>	<b>Rater 3</b>
Rater 1		<b>.758<sup>§</sup></b>	<b>.696<sup>§</sup></b>
Rater 2	<b>.758<sup>§</sup></b>		<b>.632<sup>§</sup></b>
Rater 3	.696 <sup>§</sup>	<b>.632<sup>§</sup></b>	

<sup>§</sup>Statistically significant at .03 level.

In Table 11 the average intercorrelations among the composite elements composing several second group independent variables is reported. A clear difference is in evidence between the mean intercorrelation of the five situational questions and the mean intercorrelations of the other second group multi-item index variables; the high intercorrelations of the other variables indicates the presence of a single hidden dimension of variance underlying each particular measure. Low intercorrelations between the composite rating criteria, however, are traditionally considered in a positive light as demonstrating evidence of a lack of halo error and as providing evidence that applicants are being rated on several different, independent job dimensions (Latham and Wexley, 1982).

**Table 11**  
Average Intercorrelations For Select Multi-item Index Variables for the Situational Interview Group

	<b>Sit</b>	<b>Aftersit</b>	<b>Final</b>	<b>Discussion</b>
Situational Questions	.090			
After Situational Ratings		<b>.728</b>		
Final Rating			<b>.760</b>	
Discussion Ratings				.741

Table 12 reports the correlations between each of the situational questions and the dependent variables. It is evident from this table that questions one and four have the strongest linear relationship with the job performance evaluations (performance and performance-2). Such a finding has potential for practical application in directing the performance specialist in how to hone the selection instrument to its most predictive and valid elements.

**Table 12**  
Dependent Variable Correlation Coefficients With the Five Situational Questions

<b>Variable</b>	1	2	3	4	5
On-The-Job	-.093	.064	-.038	.256	-.126
Academy	.184	-.094	.194	-.046	.166
Performance	.468 <sup>\$</sup>	-.060	.038	.332	.214
Performance-2	.418 <sup>\$</sup>	-.061	.254	.298	.296

<sup>\$</sup>Statistically significant at .05 level.

In sum, the findings from this analysis indicate that the situational interview technique is better at predicting long term job performance than at predicting success on initial job placement for correctional officers in a New Generation jail. Since the situational question ratings are job-based and do not hinge on the attribution of traits (e.g., appearance) to applicants as our comparison selection procedures did, they are less vulnerable to legal challenge. Moreover, we found that inclusion of the situational questions improves, in a cumulative sense, the ability of a selection instrument to predict effective behavior at the academy, on-the-job, and on the two subsequent performance evaluations. Finally, there is some indication that the situational questions are measuring several distinct dimensions of job behavior, creating confidence that content validity is being served

by the addition of the five situational questions to the standardized interview process.

### **Implications and Recommendations For Further Analysis**

The situational interview questions employed in this study appear to have content validity because they were created via a systematic job analysis and were judged as appropriate and related to effective job behaviors by a panel of New Generation jail correctional personnel. Furthermore, psychometric analysis of the five situational questions suggests that each question gets at a distinct dimension of performance-as would be expected from the critical incident methodology employed. In addition, analysis of criterion validity data indicates that the situational interview process, particularly the situational questions themselves, have some predictive validity over the long term. That is, high scores on the situational question ratings were correlated (.37 and .48) with high performance on the job at the June and September 1988 evaluations.

The job of the correctional officer in a New Generation Jail can be both demanding and enriching. The ability to lead, supervise and interact with inmates in the confines of a pod requires that a correctional officer possess substantial "people" skills. In order to screen for these skills among applicants, and in order to forestall legal charges of discriminatory hiring practices, the selection process should be demonstrably job-related and subject to validation. The findings from this analysis demonstrate that the situational interview instrument has the potential to meet the twin challenges of creating a validated, job-related selection device for the correctional officer position in New Generation jails. However, it is necessary that more data at different sites be collected to further document the predictive validity of situational interviews. A single location study is instructive,

but it cannot provide the level of confirmation required for broader generalizations. This analysis will also be greatly strengthened by the addition of more long term performance appraisal information to the existing data available on Newjail correctional officers.

The findings presented here are particularly timely given the recent Supreme Court ruling in *Watson v. Fort Worth Bank and Trust*, 108 S. Ct. 2777 (1988). In this unanimous opinion, the court agreed with the plaintiff that a disparate impact analysis may be applied "...to a subjective or discretionary promotion system in claims of a violation of Title VII of the 1964 Civil Rights Act" (Adams, 1988). Although the court did not go so far as to say all promotional practices must be validated to the same degree as formal selection and promotion tests, it did hold that informal selection tests such as the oft-used interview must be standardized and uniformly administered to avoid discrimination charges made on the basis of disparate impact in protected classes of applicants and/or employees. The situational interview may well represent a highly useful **middle ground** for public employers-providing some of the much needed flexibility to assess the more intangible aspects of fitness for service and-at the same time-providing enough standardization and uniformity of process to pass muster with the courts in the case of contested outcomes. At the very least, the findings reported here indicate the wisdom of further research in this area of newly immediate concern to public sector managers.

References

- Adams, J. (1988) "Legal Developments Supreme Court Decides Status of Subjective Procedures: Watson Prevails." In ***International Personnel Management Association Assessment Council News***. 12(2): 7-8.
- Agresti, A. and B. Finlay (1984) ***Statistical Methods For the Social Sciences***. San Francisco, California: Dellen Publishing Company.
- Bailey, K.D. (1982) ***Methods of Social Research***. New York: The Free Press.
- Cayer, N.J. (1987) "Managing Human Resources." In R.C. Chandler (ed.) A Centennial History of the American Administrative State. New York: The Free Press..
- Clear, T.D. and G.F. Cole (1986) ***American Corrections***. Monterey, California: Brooks/Cole Publishing Company.
- Cloward, R.A. (1968) "Social Control in the Prison." In Hazelrigg, L., eds., ***Prison Within Society***. Garden City, N.Y.: Doubleday and Company, Inc.
- Daniel, C. (1986) "Science, System, or Hunch: Alternative Approaches to Improving Employee Selection." ***Public Personnel Management***. 15( 1): 1-10.
- Davey, B.W. (1984) "Personnel Testing and the Search for Alternatives." ***Public Personnel Management***. 13(4): 361-374.
- Demarco, J.J. (1983) "Productivity and Personnel Management in Government Organization." In S.W. Hays and R.C. Kearney (eds.) *Public Personnel Administration*. New Jersey: Printice-Hall Inc.
- Dyer, F.J. (1981) "An Alternative to Validating Selection Tests." In Klingner, D.E., eds., ***Public Personnel Management***. Palo Alto: Mayfield Publishing Co.
- Etzioni, A. (1961) ***A Comparative Analysis of Complex Organizations***. New York: The Free Press of Glencoe, Inc.
- Flanagan, J.C. (1954) "The Critical Incident Technique." ***Psychological Bulletin***. 51(4): 327-358.
- Frank, E.J. Dr. (1980) "Reviewing the Obvious of Content Validation." ***Public Personnel Management***. **9(4)**: 278-281.
- Green, S.B. and T. Stutzman, (1986) "An Evaluation of Methods to Select Respondents to Structured Job-Analysis." ***Personnel Psychology***. **39(3)**: 543-564.

- Gettinger, S.H. (1984) *New Generation Jails: An Innovative Approach to an Age-Old Problem*. Washington, D.C.: National Institute of Corrections.
- Ghiselli, E.E. (1966) "The Validity of a Personnel Interview." *Personnel Psychology*. 19(4): 389-395.
- Goffman, E. (1961) *Asylums*. Garden City, N.Y.: Anchor Books.
- Guion, R.M. (1987) "Changing Views for Personnel Selection Research." *Personnel Psychology*. 40(2): 199-213.
- Hackman, J.R., and G.R. Oldham (1974) *The Job Diagnostic Survey: An Instrument for the Diagnosis of Jobs and the Evaluation of Job Redesign Projects*. Technical Report No. 4, Department of Administrative Sciences, Yale University.
- Hackman, J.R., G.R. Oldham, R. Janson and K. Purdy (1981) "A New Strategy for Job Enrichment." In Klingner, D.E., eds., *Public Personnel Management*. Palo Alto: May-field Publishing Company.
- Hays, W.L. (1973) *Statistics For the Social Sciences*. New York: Holt, Rinehart and Winston, Inc.
- Kistler, N. (1988) "The Situational Interview As A Selection Device For New Generation Jail Correctional Officers." Unpublished MA., Washington State University, August 1988.
- Latham, G.P., C.H. Fay, and L.M. Saari (1979) "The Development of Behavioral Observation Scales For Appraising the Performance of Foremen." *Personnel Psychology*. 32(2): 299-311.
- Latham, G.P., L.H.I. Saari, E.D. Pursell, and M.A. Champion (1980) "The Situational Interview." *Journal of Applied Psychology*. 65(4): 422-427.
- Latham, G.P. and K.W. Wexley (1977) "Behavioral Observation Scales For Performance Appraisal Purposes." *Personnel Psychology*. 30(2): 255-268.
- Latham, G.P. and K.N. Wesley (1982) *Increasing Productivity Through Performance Appraisal*. Reading, Mass.: Addison-Wexley Publishing Company.
- Latham, G.P., K.N. Wexley, and E.D. Pursell (1975) "Training Managers to Minimize Rating Errors in the Observation of Behavior." *Journal of Applied Psychology*. 60(5): 530-555.
- Lee, C. (1988) "Testing Makes a Comeback." *Training*. 25 (12): 51-59.

- Maas, J.B. (1965) "Patterned Scaled Expectation Interview: Reliability Studies on a New Technique." *Journal of Applied Psychology*: 49(6):431-433.
- Manning, D., Menke, B.A., Zupan, L.L., Stohr-Gillmore, M.K., Stohr-Gillmore, M.W., and N.P. Lovrich (1988) "Performance Appraisal in New Generation Jails: Model Performance Appraisal Processes and Forms and Training Materials for New Generation Jail Facilities." Final Report for the National Institute of Corrections Grant Award 87JOIGGV2.
- Maurer, S.D. and C. Fay (1988) "Effect of Situational Interviews, Conventional Structured Interviews, and Training On Interview Rating Agreement: An Experimental Analysis." *Personnel Psychology*. 41(2): 329-344.
- Mayfield, E.C. (1964) "The Selection Interview- -A Re-Evaluation of Published Research." *Personnel Psychology*. 17(3): 239-260.
- Menke, B.A., Zupan, L.L., Lovrich, N.P. and D. Manning (1986) "Model Selection Process For Podular Direct-Supervision New Generation Jails." Final Report for the National Institute of Corrections Technical Assistance Grant Award #GD-9.
- Nelson, R. (1987) "Unit Size and Inmate Management for Direct Supervision." In Wener, R. and J. Farbstein (eds.) *Proceedings of the Second Annual Symposium on New Generation Jails*. National Institute of Corrections, Jail Center, Boulder, Colorado.
- Nelson, R. (1986) "Can Cost Savings Be Achieved By Designing Jails for Direct Supervision Inmate Management?" In Farbstein, J. and R. Wener (eds.) *Proceedings of the First Annual Symposium on New Generation Jails*. National Institute of Corrections, Jail Center, Boulder, Colorado.
- O'Hara, K. and K.G. Love (1987) "Accurate Selection of Police Officers Within Small Municipalities: 'Et tu Assessment Center?'" *Public Personnel Management*. 16( 1): 9-14.
- Reid, S.T. (1976) *Crime and Criminology*. Hinsdale, Illinois: The Dryden Press.
- Robertson, D.E. (1981) "New Directions in EEO Guidelines." In Klingner, D.E., eds., *Public Personnel Management*. Palo Alto: Mayfield Publishing Company.
- Schein, E.H. (1981) "Increasing Organizational Effectiveness Through Better Human Resource Planning and Development." In D.E. Klingner (eds.) *Public Personnel Management*. Mayfield Publishing Co.
- Schmidt, F.L., B.P. Ocasio, J.M. Hillery, and J.E. Hunter (1985) "Further Within-Setting Empirical Tests of the Situational Specificity Hypothesis in Personnel Selection." *Personnel Psychology*. 38(3): 509-524.

- Silverman, S.B. and K.N. Wexley (1987) “The Selection Interview: Some Reasons for Optimism.” **Personnel Assessment Monograph**. Published by the Assessment Council of the International Personnel Management Association.
- Smith, P.C. and L.M. Kendall (1963) “Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales.” **Journal of Applied Psychology**. **47(2)**: 149-155.
- Springer, J.P. (1982) “The Importance of Selection in Public Sector Administration.” **Public Personnel Management**. **11(1)**: 9-12.
- Sproule, C.F. (1984) “Should Personnel Tests Be Used on a Pass-Fail, Grouping, or Ranking Basis?” **Public Personnel Management**. **13(d)**: 375-394.
- Stohr-Gillmore, M.K. (1987) “Inmate Perceptions of New Generation Jails.” Unpublished M.A. Thesis, Washington State University, May 1987.
- Sykes, G.M. (1958) **The Society of Captives**. Princeton, N.J.: Princeton University Press.
- Thomas, J.C. and W.D. Heisel (1983) “The Modernization of Recruitment and Selection in Local Governments.” In Hays, S.W. and R.C. Kearney, eds., **Public Personnel Administration**. New Jersey: Prentice-Hall.
- Ulrich, L. and D. Trumbo (1965) “The Selection Interview Since 1949.” **Psychological Bulletin**. **63(2)**: 100-116.
- Wagner, R. (1949) “The Employment Interview: A Critical Summary.” **Personnel Psychology**. **2(1)**: 17-46.
- Weisberg, H.F. (1974) “Models of Statistical Relationship.” **American Political Science Review**. **68(4)**: 1638-1655.
- Wernimont, P.F. and J.P. Campbell (1968) “Signs, Samples, and Criteria.” **Journal of Applied Psychology**. **52(5)**: 372-376.
- Zupan, L. “The New Generation Correctional Philosophy: An Implementation Analysis and Impact Evaluation,” Unpublished Ph.D. Dissertation, Washington State University, August 1987.
- Zupan, L.L. and B.A. Menke (1988) “Implementing Organizational Change: From Traditional to New Generation Jail Operations.” **Policy Studies Review**. **7(3)**: 615-625.
- Zupan, L.L. and M.K. Stohr-Gillmore (1988) “Doing Time in New Generation Jails: Inmate Perceptions of Gains and Losses.” **Policy Studies Review**. **7(3)**: 626-640.



Zupan, L.L., Menke, B.A. and N.P. Lovrich (1986) "Podular/Direct Supervision Detention Facilities: Challenges for Human Resource Development." In Farbstein, J. and R. Wener (eds.) ***Proceedings of the First Annual Symposium on New Generation Jails.*** National Institute of Corrections, Jail Center, Boulder, Colorado.