SELECTION THROUGH ASSESSMENT CENTERS

A TOOL FOR POLICE DEPARTMENTS

Roger Reinke

PoliceFoundation

The Police Foundation is an independent, nonprofit organization established by the Ford Foundation in 1970 and dedicated to supporting innovation and improvement in policing. This book is published by the Foundation as a service. The opinions and recommendations are those of the author, and not necessarily those of the Police Foundation.

Library of Congress Catalog Card No. 77-83720

FOREWORD

This monograph describes some of the background, most of the problems, and certain findings in support of assessment centers in the selection of police personnel. No prescriptions for remedying the ills of an assessment center are offered because the intent is to discuss possible usage within agencies possessing assessment center capability. Suggestions about establishing a center in a particular agency cannot be applied generally, without due concern for local circumstances.

Police in their police world hold sets of values that go beyond the esoteric trappings of law enforcement. If the personnel selection process does not raise these values to consciousness, it will result in a disservice to the individuals concerned and, at the least, disharmony in the profession. The assessment center method contains the means to elicit values, but it is not a panacea for predicting job success. It can, however, do a better job than many of the usual techniques.

The experience of three police assessment center users is described in this monograph, as are some of the indicators of success in the use of such a center. What is lacking, unfortunately, is a scientific analysis of results. Such an analysis will come in time, but there is already enough evidence to suggest that potential users need not wait for future analysis. Results in other fields compel attention now. In this monograph, then, the psychometrician will not find the correlations and other statistical tests he needs, but that is just as well, for this is not the best forum for such material.[1] Instead, this monograph addresses the police personnel manager and appointing authority who must live with the results of selection decisions. Those decisions are almost always judgmental. If a particular selection methodology can reduce the margin for error, recognizing that it can never be entirely eliminated, then that methodology deserves attention. Of the three assessment centers reviewed, two were the partial results of grants to the Police Departments of Rochester, New York, and Kansas City, Missouri, by the Police Foundation.

---

1. The statistical justifications for assessment center methodologies are adequately documented in the professional literature. They are not included here simply because they require, and deserve, more attention than can be given to them in this monograph.

At the time the centers were used, these agencies had, in addition to their Police Foundation support, committed money and personnel to improving criminal investigation and field patrol. The use of assessment centers fitted in well with these efforts, and the Police Foundation thus encouraged these agencies to explore the technique. The third center reviewed, in Savannah, Georgia, was used under a contract with the Professional Standards Division of the International Association of Chiefs of Police (IACP). The writer served as project director.

The author is indebted to Chief Tom Hastings and Deputy Chief John Neary of the Rochester Police Department, Captain Troy Majors of the Kansas City Police Department, and Arthur A. Mendosa, City Manager of Savannah, for their assistance. Appreciation is also extended to Richard A. Staufenberger and the Police Foundation for their interest in disseminating information to the police community about this promising selection method.

Roger W. Reinke

PREFACE

Since the Police Foundation was established in 1970, its Board of Directors and staff have devoted a large measure of time and resources to issues involving police personnel administration. This concentration on personnel reflects the fact that a significant portion of the efficiency and effectiveness of the police is linked to the selection, training, promotion and supervision of police officers.

So the general subject of police personnel has been a major program area for the Foundation and during the past several years it has sponsored demonstration and research projects in the areas of women in policing, police officer height as it relates to performance, the selection of police chiefs, psychological testing and counseling, and personnel management information systems.

So far, these projects have resulted in several Foundation publications: Policewomen on Patrol (two volumes); Women in Policing: A Manual; Police Chief Selection; Police Officer Height and Selected Aspects of Performance; Police Personnel Administration; and Kansas City Peer Review Panel.

This report marks the publication of a series of monographs on personnel issues. The subjects include performance appraisal in police departments, police selection through assessment centers, and personnel management information systems for the police.

This monograph and others in the series are published in the belief that each can help police leaders and managers in the job of improving the quality and performance of American police personnel.

Patrick V. Murphy
President
Police Foundation

CONTENTS

FOREWORD		
PREFACE		
I	ASSESSMENT CENTER METHOD	1
II	THE NEED FOR BETTER ASSESSMENT CENTER METHODOLOGIES	8
III	ASSESSMENT CENTERS IN THREE AGENCIES	16
IV	ASSESSMENT CENTER OPERATIONS	31
APPENDIX		45

I

ASSESSMENT CENTER METHOD

A Description

The task of selecting people for entry level or promotional positions has produced various techniques over the years, none of which has been strong in its predictive aspect. The traditional methods of selection in police agencies are easy to attack for their inadequacies, but it is another matter to develop an economical, ethically sound selection system that can produce better results. One attempt at finding a more reliable way to predict future performance is a method called the assessment center. As the proponents of the method constantly remind their clientele, it is a way of systematically combining various evaluative components, and not the place where it occurs.

Beyond the misleading name for the method, an important caveat sometimes is lost in detailed descriptions of the center. Police people, no different from other people trying to obtain some satisfaction out of their working lives, are also blessed with or damned by the vicissitudes of professional careers. It is unlikely that the way people regard the prospect of change and react to it will ever be fully understood or made predictable; thus the assessment center method should be seen only as a better tool in a complex setting.

The assessment center method is commonly understood to mean the structured evaluation of the relevant skills, knowledge and abilities of people by means of situational testing. The name implies the use of a variety of test exercises that elicit measurable behaviors. The tests relate directly either to the specific position to which the candidates aspire or to a general occupational level. The evaluation process requires that several observers or assessors consolidate their independent findings fairly into reasonably objective appraisals of the individual's likely performance capability. In a competitive promotional context, the assessment center results are used to compare each candidate with others in the same center, but the method is reliable to the extent that the same individual strengths and weaknesses are elicited whether or not the other participants, the assessors, or both, change.

To reduce ambiguity and thus misunderstanding about the assessment center itself, a brief description of a typical selection center is justified at this point. After a preliminary screening process identifies the people who seem most qualified for a particular position or rank, the surviving candidates are brought together. They learn that they will be observed performing various activities, then the mechanics of the center are briefly described. They are then assigned to groups of from four to six candidates each for scheduling purposes. Each observer, called an assessor, who evaluates and reports on candidate performance, carefully watches and listens to one or two candidates

during each exercise or test. These exercises may involve all candidates interacting in one group, two or three candidates working out a problem, or one candidate in a direct, intensive interview with one assessor.

Each exercise should elicit certain predetermined, job-related behavior. During and after each exercise, the assessor records his or her observations and decides upon a score for each of the observed behaviors. After all of the exercises are completed (in about two days), the assessors report their independent evaluations, behavior by behavior, and then combine their evaluations in an assessor conference. The score for each candidate is simply a tally of all the assessor scores for each behavior measured during the center. The assessors may conclude the center with a personal interview with each candidate, discussing performance and using observed behavioral examples to support suggestions for future improvement.

The exercises themselves are limited in format only by the imagination of their developer. They share common attributes, however. While they usually involve situations and problems drawn from the milieu of the position aspired to, they are always constructed to facilitate the observation of as many relevant behaviors as possible. Also, several different exercises are included to permit observations of individual behavior in varying circumstances, perhaps with a different set of schedule participants, for example.

The relationship between the position and the exercise may be obvious, as in the case of a candidate for a chief's position orally presenting justifications for the next year's police budget from supplied data. Similarly, related behaviors will be elicited by an exercise involving all group members acting as a disciplinary board and wrestling with an imaginary case, say the poor decision by a well-regarded officer to discharge his firearm. A group exercise, such as candidates acting as city council members in deciding the allocation of a windfall among the city departments they each represent, may seem unrelated to a police position, but again observations of relevant behaviors are possible. The exercise may be mostly written, as in the case of a candidate's response to a set of interrelated documents ("in-basket"), or a straightforward, probing interview about personal career goals and self-perceptions.

Regardless of the actual format of the exercise, each is intended to elicit observable, job-related behaviors in one way or another. The exercises are also designed to put the candidates on an equal footing as far as beneficial knowledge is concerned--this by using imaginary dates, times, places, events, data, and names. None of these exercises, however, should be taken as definitive; the assessment center method is not limited to them.

The assessment center does a better job of helping career development than all of the usual police selection methodologies. For example,

over one hundred centers have been conducted for career development purposes (the Professional Police Registry) by the International Association of Chiefs of Police (IACP). While the long term performance of those counseled has not been judged so far, there is substantial evidence that persons selected for promotion in their agencies have benefitted directly from assessment center results, and are performing as predicted. Moreover, evidence is matched by the impressions of those who were among the participants in a center--not an objective measure, perhaps, but certainly significant corroboration.

A Brief History

Although the use of this method by police organizations is the primary focus here, a brief review of the center's development in other fields is useful. The center's methods evolved from German military origins prior to World War I, and received much attention by the British War Office Selection Boards and the U.S. Office of Strategic Services as World War II began. The assessment experience of OSS was probably viewed by researchers in the period following the war as too specialized for general application; the situational exercises to identify potential spies did not seem germane to the selection of first line supervisors. But the method itself appealed to Dr. Douglas Bray of American Telephone and Telegraph (AT&T), who carefully developed a research design in the mid-1950s. His work has produced convincing results in the last twenty years, to the point that AT&T and its affiliates have continued and expanded its permanent role in the organization. Other organizations have not been far behind. They include industrial and service organizations, as well as private and governmental agencies, that range in size from those with full time assessment center staffs to those that employ a consultant's services as needed.

These organizations most often use the assessment center method to select supervisors and managers from the line level rather than from subordinate management levels because the center can best predict performance should the candidate move into a job with quite different duties and responsibilities. It is not difficult to predict performance if the candidate has established a record of performance and only superficial career changes are contemplated; but when he is under consideration for promotion to a new level, past performance data may become much less useful as a reliable predictive device. An analogy is the sports figure enjoying success in the minor leagues. Much data is compiled reflecting his accomplishments, but his performance does not guarantee success in the major league.

Thus far most police agencies using the center have applied it at the supervisory and mid-management level, but a center met the need as nothing else had in the case of the Mississippi Bureau of Narcotics.

The agency uses a modified center to select its entry level narcotics agents from field of diverse applicants. The FBI's Management Aptitude Program uses a center for identification of supervisory talent among agents, with further mid-management center capability under development. The list of police agencies now using the method is slowly growing; the experience of three departments is described later.

Cost and Validity

In principle assessment center methodology has merit, although the practical question of its cost makes it appear less attractive. Comparisons of "per candidate" costs in industry are not very instructive, because, for example, companies treat assessors' costs differently. Two general observations are that assessment centers for police usually cost more than conventional testing procedures, and second, beyond the remark that one gets what one pays for, it is difficult to show the cost benefits: How, for example, can one estimate the costs of "bad" selection decisions (an exercise not unlike estimating crime prevented by a preventive patrol tactic)?

Wisely, those experienced in assessment center administration emphasize the positive results. The way most users have embraced and expanded the center suggests that the costs are acceptable. In IACP's assessment center experiences, the lowest actual cost for a typical center figured out to about $300 per candidate, but obviously in-house capability can reduce this outlay substantially. Industry reports costs in a wide range, from a per candidate expenditure of $10 to $500; the most realistic costs (including assessors' and administrators' time) are in the range of $250-$400 for centers involving about two days each of exercises and compilation of assessor reports. In 1973, however, the Santa Clara County Sheriff's Department reported a $20 per candidate cost for a sergeant's assessment center. Often overlooked is the initial mandatory investment, involving: analysis of the positions to be tested for; identification of necessary skills and abilities; development of appropriate exercises; and, most importantly, training the assessors.

Assessment center validity and costs obviously require examination. Other external factors also must be evaluated: the acceptability of the center in meeting various equal employment opportunity guidelines; fair employment practices regulations, and/or federal contract compliance requirements; and the possible constraints on center usage because of civil service regulations and/or agreements with police bargaining units. The latter groups tend to share the same concerns; that is, that the competitive process be as fair and objective as possible.

Civil service agencies know of assessment center validity research, and they usually find that the procedure is objective (especially when written examinations are included as part of the center).

Police bargaining units may have agreements that specify the relative weights of certain promotional factors, such as credit for seniority. The results of the assessment center, once the bargaining unit's officers understand the method, may be treated as one of the components in the final grading; or, preferably, assessors can show the unit's officers how they evaluate experience during the center. Of course, inflexible attitudes about such "objective" factors as seniority may still remain; if they do, they can be held up as invalid predictors of future job success.

The assessment center method, in contrast to many selection techniques, does well under equal employment agency scrutiny; and in at least two cases where judicial attention focuses on police applications of the method, it was deemed acceptable. In one case arising out of allegations of unfair discrimination in entry level selection, an assessment center was not used; but the judge described the assessment center as the method of choice.[2] In another, the parties to the litigation agreed on validity of the method, but disagreed about administrative procedures.[3]

Of course, nothing prevents an agency or a consultant from calling a pale imitation the real thing. Equal employment concerns, to say nothing of the ethical concerns of any police agency in mandating a particular selection procedure, call for certain measures of non-discrimination and validity. Understanding the terminology used in describing selection procedures is important; that no longer is the exclusive domain of psychometricians.

For this discussion reliability means dependability or consistency in measurement; and validity means that the method measures what it should. Several research projects have shown the reliability of the method; similarly, its validity has been shown, but with less confidence. Confidence can be increased by demonstrating specifically, in the case of the agency using the method, that it fairly and reliably predicts future performance. If minority and female candidates are applicants, then fairness can be shown if they are successful in about the same proportion as other candidates.

For the toughest problem, proving predictive validity, the experience of others is of little value; if challenged, the burden is on the agency. Here it hardly matters whether the method looks good or the candidates feel good about it (they do).

---

2. Morrow v. Dillard, 412 F. Supp. 494, U.S.D.C., S.D. Miss. (1976).
3. Berry v. City of Omaha (D.C. Douglas County, Neb., Dock. 695, No. 31, Nov. 20, 1975).

For criterion-related validity, an acceptable correlation must exist between measurable performance standards, based on job-related indicators of performance and the original predictions. If the center's results say that a certain candidate will do a good job as a sergeant, for example, then some prior criteria must show that later a positive relationship between the prediction and the performance will exist. The question in this case is: What does a sergeant do to raise performance in this agency above the mediocre or unacceptable?

Obviously, showing criterion-related validity involves a time lag. In this example, until the sergeants selected perform or perhaps fail to perform in the real world, what needs doing is a refinement of the evaluation system against performance standards. The standards come first, then the predictors. Equal employment agencies may accept something less in the interim--called evidence of content validity. For such evidence one shows that the method entails a careful analysis of the actual skills and abilities required in the position, and that the method elicits accurate measures of them. The assessment center with its situational exercises, developed for the explicit purpose of measuring the skills and abilities needed on the job, has high content validity.

All of this suggests the need for careful planning when starting an assessment center. Of course, the same considerations apply to any form of selection procedure--written exams, oral interviews, and the like. Candidates, however, will view apprehensively the comparatively new assessment center method; and until they actually involve themselves in it, its effectiveness should be convincingly demonstrable.

This is the ideal planning point to devise ways of determining the degree of predictive validity later on. For example, in a desirable situation all the candidates for a promotional position would go through an assessment center, with the results treated in such a manner that no one in the management of the agency has access to them--not even the assessors.[4] They then evaluate the candidates at a later date using the agency's regular promotional procedure. After a number of promotions, and after carefully evaluating the performance of those promoted according to pre-existing standards, compare the assessment center predictions (now revealed for the first time) to results obtained through the usual procedure.

It may not be possible, of course, to check predictive validity for a large number of candidates. But one can at least reasonably estimate the method's accuracy in other ways. They all may be subject to what is called criterion contamination. Claims of high criterion-related

---

4. Probably no other research has been as persuasive regarding the predictive validity of assessment centers as that done by Dr. Bray and associates for their first centers at AT&T. Their research design is a model for any agency especially interested in avoiding criterion contamination.

validity fail if it appears that appointing authorities selected certain candidates for the assessment center, for example, or that supervisors rating later performance had access to (and, presumably, were influenced by) the center's results.

Validity, reliability, costs, participant acceptance, and the overall effectiveness of the process--these especially concern any potential assessment center user. Of course they concern users of other methods also.

II

THE NEED FOR BETTER SELECTION METHODOLOGIES

Those who must consistently identify people who will perform successfully need knowledge of pitfalls, introspection, and plain luck when choosing among equally attractive candidates. Many want to make these decisions, but few can claim that their predictions have always matched subsequent performance. The experienced manager of an agency or an organizational unit knows what the job requires in the way of skills and abilities, and as well as the performance patterns of those being considered for the position. Perhaps with the benefit of carefully worded and scientifically analyzed job descriptions, pages of computer printouts reporting individual performance data, and exam results, a synthesized list of eligible candidates will produce placements that result in the accomplishment of both organizational and individual goals.

Unfortunately, too many perils impede the way to this probably unrealizable state of affairs: The analyzer may fail to note important duties; the examiner may devise instruments that only measure the ability to recognize textbook answers; the evaluator of past performance--who becomes the predictor of future performance--may formulate descriptors according to nearly meaningless criteria, because of the difficulty in reducing individual police performance to a set of numbers. Policing, with an emphasis on intangible personal service, is not readily measurable into units of production. The point is simply that the process of selecting a few from among many calls for exceptional judgment. Assumptions made at every stage, because nothing better exists at the moment, always threaten the soundness of that judgment.

Whether articulated or merely expressed as resignation to the system, those involved generally see the process of personnel selection as lacking both efficacy and integrity. Some managers simply rely on common sense, making subjective judgments rationalized on the basis that previous selection decisions seemed to have worked out well. That may or may not be; marginal performance may be tolerated for long periods of time. In the typical, closed-to-outsiders police selection process, unsuccessful aspirants swallow hard and resume their tasks, patiently hoping that the failures of the appointed will become too obvious to tolerate, or that they themselves can somehow achieve favored status, or that vacancies will occur through retirement or misfortune.

Because civil service systems seek to establish and maintain selection processes that reward merit, efficacy and integrity are implied.

Too often, however, the system serves its own bureaucratic proclivities, and may simply operate as pro forma sanction for the decisions of others. The final measure, after all, is in the performance of those selected. Inspection of the record suggests that there must be better selection procedures.

Of course, one can point to notable improvements, if not innovations, in the art of personnel selection. Putting aside the informal selection procedures of non-merit systems in which one's political party affiliation and involvement count, the usual formalized procedures at least approximate a system, and may give a reasonable number of candidates some opportunity to compete on an equal basis. But consider the following, which could well gnaw at the decisionmaker: Ten thousandths of a percentage point separates in strict rank order a large number of candidates; but some move to the top of a competitive list simply by component weightings (e.g., veteran's preference) that may give certain questionable factors unintended influence.

Probably the most critical of selection decisions occurs for the first level of supervision; success here directly and obviously affects the daily operations of the agency. There should be little doubt that police agencies in general do not consistently identify capable supervisors. Unfortunately, once that first promotional level is attained, subsequent promotion often appears as the lesser of perceived evils--demotion or continued ineptitude. The well-worn admonition that the best police officers do not necessarily make the best supervisors is occasionally ignored, with resulting damage to both the individual and the department. Since most police administrators patiently follow the traditional career ladder, it seems that many more have acheived leadership positions despite the system than because of it.

The quest for objectivity in the preparation and grading of testing instruments has led to emphasis on written exams that permit only one correct answer, at least in theory, on subject matter deemed important to the position. The greater the competition for comparatively few promotional positions, the greater the emphasis on written exam results. Hence, the selection agency may focus on assuring a "good" distribution of test scores rather than on relating test items to actual job requirements.

Recognizing that not all police supervisory and management skills are evaluated well through written exams alone, police managers place greater weight on oral interviews as the rank aspired to increases. But apprehension about oral interviews exists, ranging from questions of structured versus unstructured interviews to participation of citizens versus "professional." As with written exams, the inherent limitations of the process (in the sense of evaluating actual job behaviors) are often glossed over.

Economics encourage simplifying the selection process to the greatest extent possible, especially in situations with a large number of potential competitors. Police chief opportunities, for example, open to applicants outside the agency, may typically result in about 100 competitors.[5] Large city sergeant exams may produce hundreds of applications, from almost all officers with a few years' service. The use of batteries of tests to produce a panorama of measures is impossible, and thus managers reluctantly decide according to what is readily available, with no developmental expenditures required.

Most merit systems provide an escape clause in this well-intended procedure, known as a probationary period. Although more often found at the entrance level, mid- and upper-level positions may require new appointees to demonstrate their ability to perform before securing permanent status. However, the number of probationary appointees actually returned to their original rank is a very small proportion of the total appointed. While this seems to prove the ability of the system to identify the top performers, it in fact only confirms that no administrator (or bureaucracy) likes to admit errors of judgment.

The search for better selection methods has been accelerated by outside pressures (for example, the federal government's requirements for equal employment opportunity and, at times, police associations or unions bargaining for "objective" selection criteria) as well as by general discontent with selection methodologies. There is little agreement, however, about the most appropriate selection methodology to pursue for various levels or even within each particular level.

During the period of 1973-75, an informal follow-up questionnaire was sent by the IACP Professional Police Registry to each appointing authority that placed a job announcement in The Police Chief. The results of this informal survey apply only to chief of police openings, since that was by far the most frequently published opportunity and the only rank for which results were tabulated. Also, the fact of publication in a national journal implies the positions were open to outside applicants, and that has some influence over selection procedures. The occasionally expressed opinion of unsuccessful outside candidates that their participation merely makes an internal selection more attractive was not verified in about three quarters of the surveyed appointments. The insider, of course, enjoys a decided advantage most of the time; yet, the comparatively low success rate for insiders in the 95 surveyed cases tends to contradict the outsider's assertion.

Written exams were used only by 22 per cent of the agencies, a small proportion contrasted to the 71 per cent that employed a formal oral interview procedure. Every agency using written exams also used oral interviews.

---

5. A vacancy in the chief's position in Lighthouse Point, Florida, attracted 352 applications.

Personal interviews by the appointing authority (in addition to a more formal oral board procedure) were used by almost all agencies, but 41 percent of the agencies assigned specific weights to the personal interview. From comments and descriptions contained in the questionnaires, the written exam was used mainly to reduce the number of candidates rather than as a specifically weighted component in the total process.

The number of applications received in response to nationally advertised positions creates a substantial problem in identifying those most qualifed. Some of the techinques used by appointing authorities have been described in Police Chief Selection.[6] In the 95 agencies described above, located in 28 states, resume review by a selection committee and/or the manager or mayor was the straightforward method of choice for most respondents. Typically, about 100 applications were received, with about 75 deemed worthy of further screening.

For the traditional methods of written exam/oral interview, little agreement seems to exist about the appropriate weights the various components should bear. For the oral interview board, for example, 14 of the 67 agencies used only internal staffing; the rest relied upon outsiders or a combination. A conclusion that may be drawn from this survey is that while many rely on oral interviews to identify leadership, few agree about their importance or the best way to conduct them. Probably no oral board interview procedure will neatly fit the needs of all agencies.

The regulations of civil service departments also influence selection methodology. Undoubtedly, some of the weights assigned to written exams/oral interviews used by the agencies responding to the survey above were specified by regulation. The rationale for determining these weights involves approximating the comparative value of each component in eliciting measurable job-related skills. Commonly, the specified proportions of each candidate's score on each component are added to produce a rank-ordering based on a perfect score of 100 percent. In reality, of course, unfair rank-ordering of candidates may occur unless the standard deviations of all components are identical.[7] Thus, standardized scoring is called for, but may prove too bothersome to calculate or, more often, cause misapprehensions about the accuracy of final results because of the unfamiliar numbers produced.

---

6. Michael Kelley, Police Chief Selection:A Handbook for Local Government (Washington, D.C.: Police Foundation and International City Management Association, 1975).

7. Terry Eisenberg and Roger W. Reinke, "The Weighting of Scores in Promotional Exams," The Police Chief June 1972, 46 ff.

An occasional source of frustration in seeking improved methodologies is the granting of grade points to veterans, usually after they have attained a specified passing score. An increase in the number of agencies employing veteran's preference was noted from 1970 to 1972 in a 1973 survey;[8] but anecdotal evidence at least suggests considerable dissatisfaction with the practice as the number of veterans seeking police employment diminishes (and, perhaps, as selection methods are improved, especially in terms of job-relatedness).

As noted, oral interviews play an important part in almost all competitive situations, and especially at upper levels. Civil service systems have promulgated many techniques to reduce or control inevitable subjectivity. But the persistent opinion of personnel managers is that a number of candidates in every situation have the quality of "presence," resulting in the awarding of artificially high scores. Their opinion has some basis: Reserach into oral board scoring tendencies shows disparate effects of "presence" and several other factors on final scores. These include not only the halo effect of quick rapport alluded to here, but also the candidate's place in the list of interviewees, his final impression, the training of interviewers, scoring procedures, and the like.

The concern is not that the candidate can successfully project an image of confidence and command ability, which may not even be a job requirement, but that the display may be superficial, hence not accurately evaluated in the oral environment--even by experience interviewers. The same dissatisfaction is often expressed by candidates who know their fellow competitors, and feel that they will impress the board to an extent unwarranted by their actual abilities. The reliance placed on oral interviews is probably even more extensive in other fields than in the police service. While continually stating the need for more research, industrial psychologists seem to feel that the oral interview technique is reasonably valid under certain conditions. Most police applications pay little heed to those conditions, however.

The weaknesses noted in existing police selection methodology can be overcome at least partially by concentrating research and experimentation activities on each specific phase, and gradually evolving procedures that are demonstrable improvements. This kind of attack tends to perpetuate the same systematic weaknesses, however; better written tests, more objective oral interviews, and the like are the goal, but the result may not yield a concomitant increase in the number of candidates who live up to expectations. For example, the test may efficiently measure skills in a discriminating way, have high content validity, and have no adverse impact on minorities; but the decisionmaker who mistrusts such tests may largely ignore them.

---

8. T. Eisenberg, D.A. Kent, and C.R. Wall, Police Personnel Practices in State and Local Governments (Washington, D.C.: Police Foundation, 1973).

Incremental improvements, although helpful, do not satisfy the lingering hope for better selection procedures. The assessment center employs some or all of the elements that compose most current selection procedures, but in ways that enhance consistency and objectivity. Police personnel managers are not alone in their need to improve selection procedures, but the characteristics of police careers especially require procedures that can withstand a variety of challenges not often found in the private sector.

The usually tedious work involved in determining sensible measures of job performance and showing job relatedness should lead to a review of other personnel programs. Given identified performance predictors, and some working knowledge about the center, then the assessment center method will complement career development systems and other special applications. (If all the assessment center operations in both private business and public service were reviewed, many more would be found for career development/potential than for straightforward selection purposes.) Certainly, the concerns of validity and job-relatedness apply also to the task of preparing individual career development plans.

The original assessment center use by the Mississippi Bureau of Narcotics in 1973 was for two purposes: providing additional information for making the first promotions to the supervisory rank (the Bureau in its brief history had made no promotions up to that time), and especially helping to identify the organizational assignments to best match individual agency needs. In the three years since the original centers were conducted, the predictions have proven to be highly accurate; also the involvement of personnel in first receiving assessment center feedback about their own skills and abilities and then becoming responsible for subordinate evaluation has made them unusually sensitive to the many problems in career development. Moreover, assessment center consultants claim that the training to become an assessor provides a better understanding of performance evaluation in general. The experienced assessor can identify the sometimes subtle behavioral roots of performance, and likely will not do violence to the department's rating system, ill conceived though it may be. Finally, the center results contain many indicators of overall training strengths and weaknesses.

There are two uncommon situations in which the center can be especially useful. Most police agencies do not recognize specialists by formal rank, and usually reward specialists' skills by grade pay differentials, but little else. A few agencies attempting to gain greater flexibility in personnel management have added specialist classifications such as police agent to the rank structure, primarily to attract applicants who would not ordinarily seek a police career, or to offer greater incentive to current members of the agency who, for whatever reason, do not aspire to higher ranks through traditional promotional procedures. When vacancies are announced, a large number of candidates may offer euqally impressive college credentials or supervisory recommendations in seeking the position.

An assessment center in this situation may specifically identify superior specialist skills as well as general promotional potential.

The other noteworthy application occurs in the comparatively rare situation involving competitors for a position, usually of chief, from both within and outside the agency. Typically, the outside candidates feel that the insiders have unfair advantages, and the insiders feel threatened by the prospect of an unknown outsider coming in. The loss of a promotional opportunity regarded as rightfully theirs adds further to insider disgruntlement. With conventional selection procedures, the candidates seldom, if ever, get a chance to know their fellow competitors, and the final selection decision seems suspect to them. The assessment center brings the candidates together in both a formal way during group exercises, for specific interaction with their peers, and informally during breaks such as for coffee and meals. This interaction usually means more knowledge of and appreciation for the abilities of the eventual appointee, and less speculation about extraneous influences in the selection process. Of course, that knowledge can make the task of the appointing authority a little more difficult, in the sense that even more people will be looking over his shoulder.

The discussion above of assessment center advantages needs balancing by notice of possible problems. The center's role as a component in a larger evaluative setting has been recognized, as has the higher costs usually associated with the method because of assessor training, exercise development, actual administrative time, and so forth. Most of the participating candidates do not care about the exercises themselves, their programming, or the behaviors evaluated. They care about who assesses them, the sometimes arduous schedule involved, and the lack of a final score that conforms with or is similar to prior scoring procedures. If these concerns go unaddressed, candidates may see the whole process as faulty. A lesser problem, unless a police knowledge exam is given, is the absence of study material. Candidates apparently feel uncomfortable about the prospect of a highly competitive procedure without tangible preparation for it.

The most frequent problem in the design of a police assessment center is determining the best way to reduce the number of candidates to workable size. The methods used by three police departments for this will be described below.

To summarize, the assessment center can be used in a variety of selection situations. It should be regarded as the method of choice when the number of candidates is small, assuming limited resources, and at that at least some of the following circumstances apply:

- When the position or level selected requires skills and abilities different from those used by candidates in their present jobs.
- When following preliminary review, the candidates <u>appear</u> equally qualified, in their own ways, for the position or level.

- When the final group of candidates consists of those within and from outside the agency.
- When there is a need to report and interpret results beyond mere rank ordering of candidates.
- When there is adequate time to analyze the position, establish performance criteria, plan and develop materials, and train assessors.
- When, without regard to formal selection for promotion, a formal career development program is desired.
- When employee organizations as well as upper level managers dislike conventional selection methods, and show a willingness to try a new approach.

III

ASSESSMENT CENTERS IN THREE AGENCIES

In police circles, the assessment center method has been tried only in a few agencies to date. The New York Police Department started its center operations for upper level management ranks and then extended them to lower ranks. The Mississippi Bureau of Narcotics uses certain assessment center techniques for entry level selection. Other departments with impressive internal programs are the Kansas City, Missouri, and Rochester, New York, Police Departments. Describing how these latter two agencies applied the method, the attitudes of personnel toward it, and its results may help others seeking solutions to their own selection problems. Finally, the experience of an agency (Savannah, Georgia) making extensive use of an outside consultant will be described.

Kansas City, Mo. In Kansas City, the eligibility lists for sergeant and captain (the rank of lieutenant is not used) expire after two years. In late 1974, the time for a new sergeant's exam was rapidly approaching. Past practice was to administer a written exam, with oral interviews conducted by outside, mostly non-law enforcement people for those who passed the written portion. The method was not well regarded by candidates, primarily because of the perceived likelihood of uncontrolled biases affecting the findings of the interview boards.

Following a visit to New York by Kansas City representatives, the department's personnel manager advocated the use of an assessment center for the upcoming sergeant's exam. Fortunately, the department retains almost exclusive jurisdiction over personnel management functions, and thus did not need to persuade a city or state civil service agency to embrace the plan. The top administrators of the department approved the concept, and the Police Foundation provided some financial support, since improved supervisory selection would complement its other projects in the city. A general order was issued describing a three-phase selection: a written exam weighted at 25 percent of the final score; a candidate review board, to screen out candidates with adverse disciplinary records, disqualifying physical conditions that were not sufficient in themselves to justify dismissal, or emotional stability problems viewed in a like manner; and an assessment center phase weighted at 75 percent.

The reaction of the prospective candidates was favorable. They felt that the lack of appreciation for the law enforcement role by the former oral boards would be remedied. As candidates studied departmental regulations and two basic police management texts for the written exam, current supervisors helped develop statements describing as specifically as possible their perception of the tasks sergeants perform.

A consultant developed a content-valid written exam, and another consultant was employed for the task analysis and assessment center work.

Several conferences were held with sergeants to gain consensus about the most important tasks required of sergeants. Using this composite task analysis, the sergeants and the consultant matched tasks with the skills and abilities needed to perform them. The latter were taken from a commercially available list of representative personal traits, or, as they were described in the context of an assessment center application, behaviors. They are also called dimensions. By whatever name, they help apply rather general descriptors, such as "leadership" or "decisionmaking," to the particular attributes of the position under consideration.

The Kansas City sergeants agreed that about 14 dimensions conveyed the most important traits they felt sergeants must possess. Because some dimensions seemed more important to job performance than others, a group of 40 sergeants with extensive field experience were asked to rate the relative importance of each dimension on a scale of zero to 100. Their combined judgments produced nine dimensions, each of which was assigned an appropriate weighting factor. Subsequently, during the training of assessors and in the assessment centers themselves, the weight of each dimension was not revealed to the assessors or candidates. The dimensions were always listed alphabetically on the various forms used to report assessor observations, and always with the definition of the dimension, to convey the fact that all dimensions were of some importance and to enhance common understanding of the meaning of dimensions.

The assessment center exercises were selected with the help of the consultant and constructed to display the nine dimensions in as many exercises as possible. The general order announcing the new procedure indicated that the top 125 scorers on the written exam (some flexibility was retained to choose a point where an obvious gap in scores occurred) would proceed to the Candidate Review Committee. This cutoff was accepted by candidates, because it ended the practice of relying solely on an arbitrary percentage of correct written exam answers to determine if a candidate passed or failed. The number of sergeant vacancies anticipated during the two-year life of the new list would come nowhere near 125, so the cutoff generously favored the candidates.

As the task analysis work proceeded, with the field sergeants deeply involved in developing the nine dimensions, the quality of credibility was established. That sergeants were included in important ways in the new process impressed the line personnel. Further acceptance occurred when personnel learned that three trained departmental assessors would determine scores, as opposed to the potentially adverse influence of one supervisor. Thus, the potential competitors at this stage commented positively.

The Candidate Review Committee did its work without seeming like a hatchet agency for the administration, primarily because of the small number actually rejected, and the nature of its decisions, which were

rather obvious to those who had some knowledge of the individuals and the decision factors. Also, the factors and procedures were fully described in the exam announcement.

Step-by-step descriptions of the assessment center procedures in Kansas City, and in each of the other agencies noted in this chapter, appear on Chart I. Starting with the initial screening process, the chart follows and compares center development through the types of exercises used in each, and the number of candidates involved in each phase.

Because the process of selecting, developing, and scoring exercises was identical in each of the center applications described, discussion of these aspects as well as actual center supervision will be found in Chapter IV.

In Kansas City, the next step was to select and train the assessors. The personnel manager, sensitive to the problem of who should evaluate candidate performance, decided that current captains, especially those regarded as particularly competent in developing the skills of their subordinates, and who were receptive to specialized training, offered the best promise. They had performed well as sergeants; they certainly knew the department and what police work was all about. The likelihood of personal favoritism on the part of internal assessors did not appear to concern the departmental administration as much as did demands on middle management time. The personnel manager correctly anticipated that the use of captains (again, the rank of lieutenant is not used) would tend to improve overall performance rating skills, and that the use of internal assessors would more than compensate for some apprehension about the nebulous problem of favoritism--which the manager felt an assessment center could control. These decisions, to make extensive use of sergeants in identifying job-related skills and abilities and in assigning a scale of importance to them, and to use internal assessors from the next highest rank, contributed greatly to line level acceptance.

The actual training of assessors took place over a two-day period, starting with background about assessment center methodology, and continuing with dimension definitions, use of reporting forms, scheduling and scoring requirements, and, finally, work with staff people acting as candidates participating in each of the various exercises.

Two days for assessor training probably places an unfair burden on the conscientious novice assessor. He or she must master dimensions and their definitions, around which the whole center is constructed, in a few hours. The assessor must learn to recognize the visual and aural clues that reveal germane behaviors, while simultaneously noting examples for later discussion. In Kansas City, that the assessors went through practice sessions, after which a critique was offered, probably saved the day.

Following assessor training, a consultant-recommended ratio of one assessor for every two candidates became the basis for devising an 18 assessors, 36 candidates per week schedule. This arrangement, over a three-week period, seemed to make minimal demands on the assessor captains' time. Scheduling flexibility was also gained by rotating the assessors among the three-man assessor teams.

The critical dimension scoring procedures were monitored closely, and later analysis showed that the scores awarded by each assessor were consistent with those of the other assessors. One assessor, however, seemed slightly more generous than the others, but the practical effect of this was minimal, because of the nature of the assessor scoring conference in which assessors report and discuss their observations. Any pronounced disparities in scores must be worked out. Compensation for the assessor who tended to give high scores undoubtedly occurred in the scores of the other participating assessors. In a later assessment center application in Kansas City, keeping the assessors on the same teams throughout the center produced at least a questionable scoring pattern for one team.

Following the last center and publication of the candidates' positions on the eligibility list, a week-long supervisory training course was given to the successful candidates. An exam at the end of the course simply qualifed the candidates--no one failed. The exam was not part of the actual assessment center.

After completion of the promotional process, those who participated either as candidates or as assessors maintained their positive attitudes. One candidate out of the 117 challenged his score, but not the methodology. A preliminary statistical analysis revealed no adverse minority group impact; also, age, tenure, and written exam scores did not influence assessment center scores. Discussions with a candidate about his or her performance were offered on a voluntary basis, and 63 candidates (about half of the group) accepted the opportunity.

At least one assessor felt that the quality of the feedback information was not as high or as consistent as it could have been, because the emphasis at the time of assessor training was on scoring for rank order purposes. Sufficient time for assessors to develop useful and convincing feedback material was not available, in his opinion.

One of the persistent questions about assessment center methodology concerns exposure to the process, either as a former candidate or as a previous assessor. In Kansas City, after the sergeant's center had been run, an investigator's center was conducted along the same lines. Several candidates who had competed in the sergeant's center were evaluated in the investigator's center. While a statistical analysis has not determined the degree of performance correlation, strong anecdotal evidence suggests that candidates did not benefit unduly from their better understanding of the process. One assessor who had also helped to develop exercises for the sergeant's center felt that his performance

in a subsequent captain's center was actually hindered by his "insider's" knowledge.

Rochester, N.Y. The Rochester application of the center methodology was for the investigator's position, and complemented extensive management attempts, undertaken by the department with Police Foundation support, to improve the investigative function. As in Kansas City, dissatisfaction with the prior appointment procedure for investigators, in this case heavily influenced by political considerations, stimulated the quest for better selection criteria. The police bargaining unit was deliberately involved in the early discussions about a better procedure, which led to a formal agreement that "merit testing" would replace the prior practice. The assessment center method was favored, largely due to the observations made of the Kansas City application; that the Chief of Police and another top staff executive had participated in assessment center surely contributed to the decision, as did the inspection of the Kansas City center by the bargaining unit representative.

Rather than rely on a written exam to reduce candidates to a number that could reasonably be scheduled through an assessment center, the department elected to use performance evaluation to identify the most promising. Appropriate caution was exercised here. The same reasons a written exam was regarded as unsatisfactory in the measurement of investigative potential could well apply to a performance evaluation process that failed to address the skills and abilities related to the investigator's job. The first task, therefore, was to undertake a job analysis of the investigative positions, which would not only serve as a foundation for assessment center exercises, but also for the performance evaluation system. The County Civil Service Commission, having no specific responsibility for investigator selection, agreed that the position should remain exempt.

The consultant employed to perform the job analysis was particularly successful in generating useful information by training 20 investigators, selected through consensus among their supervisors, to perform analysis work. A 40-hour course in analysis techniques was built around a model of the investigative process, represented in a flow chart devised by two investigators who served as project administrators. The consultant guided and reviewed the week-long work of the analysts. The product was a set of functional criteria, each criterion defined in terms of component tasks and arrayed on a scale.

The criteria and their definitions formed the basis for a task analysis report: a listing of the skills, knowledge and abilities required to perform the Rochester investigative functions, and a job description. For example, selection criteria for the Rochester investigator's position includes "people-oriented criteria" as a part of "functional requirements":

- "The applicant can demonstrate the ability to befriend and encourage individuals on a personal, caring basis either in

one-to-one or small group situations; give instruction, advice, and personal assistance concerning activities of daily living and the procedures of various institutional programs."

- "The applicant can demonstrate the ability to persuade or influence others in favor of a point of view or course of action. (Competence in relation to criteria can be demonstrated in job-specific exercises or tasks, or in exercises drawn from common life experience.)"

A few of the many adaptive skill requirements follow.

- "Ability to remain task-oriented while coping with distractions/diversions."

- "Ability to communicate with a wide variety of people (adapting language to subject's level and culture; tolerance for diversity among people; no appearance of superior/inferior attitude toward people of different race, ethnic or cultural background, sex, and social class)."

- "Ability to remain continuously alert to the unexpected and to internal signals of what 'rings true' and what does not, what 'fits' and what does not. . . ."

- "Genuine curiosity and interest in people. (This seems to be necessary for one to maintain the keen and continuous observation required in investigative tasks and may be the base which allows the detective to integrate the complex sociological and psychological knowledge necessary to recognize different motives, patterns of behavior, character types, personalities, etc.)"

- "Ability to keep several complex images in mind until thorough comparisons among them can be made (managing the stress of complex mental work dealing with several abstract and concrete variables simultaneously through time)."

A superficial view of this meticulous work might conclude that the detail in the analysis was unnecessary for the criteria needed for exercise design; however, a solid basis was created, not only for selection purposes, but also for performance evaluation and investigative training. The investment was well justified. A byproduct of the task analysis training was the emergence of six of the 20 investigators as especially proficient in task analysis, and they will be used in future updating.

The task analysis report was forwarded to the consultant charged with selection development. This consultant was the same firm used for assessment center development in Kansas City. Using the description of the required investigative skills and abilities produced by the first consultant, 12 behavioral dimensions were developed that contained the skills and abilities amenable to objective measurement in an assessment center. A few abilities, such as proficiency in the use of firearms, were properly identified as entry level requirements, and thus not an

appropriate target for assessment center evaluation. The dimensions were then weighted by soliciting the opinions of past and present investigators. The relative assigned weights of the dimensions were not closely guarded in this case.

One of the dimensions common to both the Kansas City and Rochester task analyses was that of "risk-taking" ("the ability to take calculated risks based on sound judgment"). In both cities, the dimension was removed from assessment center evaluation because of the inability of assessors to distinguish consistently between purely intellectual risk-taking and physical risk-taking.

Information about the candidates' physical, emotional, or disciplinary problems was available before the start of the first major step, the performance evaluation; but, because of appeals, the panel actually was operating concurrently with the evaluation phase. A panel of all department section commanding officers (generally with the rank of captain) reviewed information in department records as well as the responses submitted by candidates whose fitness was questioned by the panel. No candidate could be removed from competition without first having had an opportunity to be heard by the panel. Of the seven candidates so affected, four withdrew from further participation, exercising an option designed to prevent personal embarrassment, two were rejected, and one was permitted to continue.

The performance evaluation phase caused the most concern among candidates. The 37 definitions of skills and abilities reported in the task analysis were translated into six behaviorally anchored scales, each with four behavioral examples (these six scales, in effect, were the six most important dimensions, in the consultant's opinion).

Sergeants, co-workers, and peers were chosen to rate performance, and lieutenants were selected to review the raters' judgments. Two sergeants were selected to rate each candidate on the basis of at least six months' association within the year. Seven peers who met the six month test for knowledge about each candidate were identified, and then two were randomly selected from the seven. The lieutenants, called facilitators, interviewed the raters, asking for behavioral examples to justify the reported dimension scores on the six scales. The facilitator could not change a rater's score, but did seek additional examples, if needed. Facilitators were assigned raters from different organizational units. In this manner, it was felt that subjectivity could be adequately controlled. The completed performance evaluation reports identified the 150 top scorers (plus nine tied for the 150th position). None of the performance evaluation scores was released to the assessors.

The Department initially intended to use performance evaluation to reduce the field of candidates to workable size for the assessment center phase. There appeared at this time, however, a number of reports from candidates alleging that the procedure was too subjective and unfair.

CHART I

Assessment Center (A/C) Selection Steps in Three Departments

	Kansas City	Rochester	Savannah
Rank	Sergeant	Investigator	Chief
Screening Step I	Written exam (150 items, 3 hours)	Candidate Fitness Panel	Resume
Conducted by	Exam Consultant	Section Commanders	Review by consultant and appointing authority
Number of Candidates	469	240	112
Number Surviving	125	235*	17
Screening Step II	Candidate Review Committee	Performance Evaluation	Verification of intent and credentials
Conducted by	Captains and above	2 Supervisors and 2 peers per candidate (interviews conducted by trained facilitators) (Lts.)	Consultant
Number Surviving	117	159	12
Screening Step III	None	Case file exercise (in-basket)	None
Conducted by		Assessors (Sgts.)	
Number for A/C	117	88**	11**

	Kansas City	Rochester	Savannah
Task Analysis			
Step I	Questionnaires to sergeants	Consultant trained investigators	Review of prior task analysis work
Step II	Interviews with sergeants	Interviews with investigators	Conference with appointing authority
Step III	Consultant guided conference with 7 sergeants	Task consultant report to A/C consultant	
Number of Dimensions Derived	9	11	22
Weighting by	40 sergeants	101 investigators	Consultant
Weighting Known to Assessors	No	No (revealed after assessment center)	No
Assessor's Rank	Captains (next highest rank)	Sergeants with investigative experience	Consultant's staff and chiefs
Assessor/ candidate ratio	1:2	1:2	1:2
A/C Exercises	3	4	6
Group discussion I (candidates assigned roles)	-	"Investigator Advisory Committee"	"City Council"
Group discussion II (unassigned roles)	"Management Problems"	-	"Shooting Board"
In-Basket	30 items, 3 hours	30 items, 2.5 hours	22 items, 2.75 hours

	Kansas City	Rochester	Savannah
Fact Finding	-	1 hour	-
Career Interview	1 hour	1 hour	1 hour
Written exam	Davis Reading Test (1-D)***	Davis Reading Test (1-D)***	40 items, 1 hour
Oral Presentation	-	-	7 minutes each

*Panel questioned seven, of whom four withdrew, two were rejected, and one was permitted to continue after interview with panel. Subsequently, one was disqualified by A/C administrator for less than required two years of service.

**One candidate voluntarily withdrew immediately prior to the center, and was replaced by next ranking case file exercise participant.

***The results were not included in the scoring procedure.

The decision was made to minimize the importance of the performance evaluation phase by using a written test built around a homicide report. The test was named the case file exercise, and was originally intended as part of the actual assessment center, rather than as a pre-center screening device. It was the equivalent of the usual assessment center in-basket, an exercise requiring the candidate to coordinate, plan, and implement (in writing) decisions dealing with various job-related situations.

The candidates' reaction to the case file exercise seemed to counter the adverse feedback about performance evaluation. The exercise had high content validity, being a synthesis of 15 factual major cases successfully concluded by investigation in Rochester (except that ficitional names and so forth were used). After the test was administered to the candidates, copies of the work were given to each of the assessors for independent evaluation; the original work of the candidates was sealed. An interview to asertain the candidate's rationale for his decisions was scheduled usually within the three days following the exercise. The team of two assessors handled three one-hour interviews per day, taking an additional hour for each to report scores and behavioral examples for the dimensions elicited in the exercise.

Twelve assessors and three administrators were designated as the assessment center phase began. Assessor training had been completed at this point, similar to the Kansas City procedure in approach, but three days were provided. The credibility problem was helped by the stipulation that the assessors, all sergeants, had to have investigative experience, but some candidates expressed reservations about the quality of that experience in a few cases. The 88 surviving candidates were scheduled for six-person centers over a four-week period, with the assessors rotated among the teams of three persons.

The latter plan caused some logistical problems because of a thoughtful proviso that a candidate could disqualify any one assessor, much as a pre-emptory challenge, and similarly the assessors were instructed to disqualify themselves should they have a candidate with whom they had worked or had regular social contact, or whom they otherwise declined to evaluate. The project director devised a schedule taking these challenges into account, while dispersing the candidates through the schedule so that placement would not reflect the scores attained on the case file exercise.

A natural concern was the extent of contamination arising out of informal discussion about the procedures as candidates were processed. There was no obvious "inside" knowledge for the candidates, however, except for the case file exercise. The delay between the time the exercise was first administered and the subsequent interview (as much as three weeks) worried some assessors: Candidates might compare notes about how they handled various items.

Interviews with both assessors and candidates revealed an acceptance, even among close friends, that in a competitive situation such as this, sharing useful insights would only tend to reduce chances of success. Later, in the actual assessment center, several candidates noted afterwards that because they were well known to other candidates in group exercises, deliberately faking behavior would be disadvantageous.

The scoring scheme was to weight performance evaluation at 25 percent and assessment center performance at 75 percent, as in Kansas City. The case file exercise score determined the highest scoring candidates who would be admitted to the assessment center; the actual case file exercise score of the surviving candidates was then included in the assessment center score.

Despite the proportionately low impact of the performance evaluation results, candidates regarded it as the weakest component. Rumors circulated that some deals occurred among the peer raters, and some candidates thought the evaluation was too heavily influenced by present assignment, especially if the candidate were in a small, cohesive unit such as vice control, internal investigations, and the like. Some of these feelings were based on the past history of political interference in the promotion and assignment of investigators. Evidence of rater bias, noted in a few appealed cases, was resolved, however. The department plans in the future to require the rater, rather than the facilitator, to actually note performance evaluation scores, and to justify scores with recorded behavioral examples.

While the performance evaluation problems were bothersome, both assessors and candidates liked the overall procedure. About 30 investigators were appointed as a result of their performance, and interviews with competitors who finished at both the top and bottom of the list indicated a general feeling that the "right" people were in the upper range in reasonably accurate order.

Candidates in both Kansas City and Rochester, besides sharing a favorable opinion about assessment centers generally, seemed to have similar reactions about the center exercises. Although Rochester candidates knew about the dimensions, that was of little help in understanding why a variety of exercises was used. In both places, the in-basket or case file exercise was viewed most favorably (a "fact finding" exercise in Rochester was close behind), and a career interview exercise was held in least favor (too susceptible to the whims of one assessor). The group discussion exercises were given middling approval. One candidate was annoyed that an assessor during the career interview told him he had performed very well in the group discussion exercise, only to have that report contradicted in a later feedback interview. Understandably, he questioned the reliability of all the feedback suggestions as well as his score on the exercise.

Savannah, Ga. In the summer of 1975, the city manager of Savannah faced the prospect of appointing a new chief of police, following the retirement of the incumbent. The department had accomplished some notable improvements in the delivery of police services, and in stimulating the pursuit of college level training by all ranks.

The manager could reasonably expect competent leadership if he chose from among the members of the department, but hiring an outside administrator looked advantageous because he or she would not have to deal with all the relationships, some good but others adverse, that inevitably come with a long career in one agency. Of course, the alternative had its drawbacks: The outsider would have to deal with speculation about his style and policies, as well as the antagonism coming from denial of the chief's job to some members who worked hard over the years in hopes of it.

The manager's assessment of the situation led him to seek a process that would fairly evaluate the potential of both inside and outside candidates, because he recognized that a competent police leader would prove his abilities regardless of previous affiliation. Learning of the assessment centers used in like circumstances, the manager evaluated the results in other cities and began negotiations for the services of an assessment center consultant. An announcement of the opportunity encouraging applications without restrictions as to present rank was circulated in the department and in The Police Chief magazine, as well as elsewhere. Resumes were solicited from those who met rather liberal experience prerequisites ("...law enforcement experience at the administrative or management level"). A college degree was preferred, but not necessary.

Predictably, the manager received some pressure from the community to appoint a current member of the department. At the deadline for applications, among the 112 received, there were several well-qualifed candidates from within and from outside who warranted further consideration.

Resume review was undertaken independently by both the manager and the consultant, who then combined their independent evaluations. The procedure used by the consultant was first to eliminate those who did not meet the experience requirement--applicants who showed only first line supervisory responsibilities or less. The manager's approach was based on his evaluation of potential related to resolving community problems and the candidates' independent ability to develop and implement programs. Those candidates whose resumes indicated that they recognized the value of acquiring a broad knowledge about police operations and management, and who demonstrated the ability to apply what they had acquired, were jointly selected for the actual assessment center.

The deadline for filing applications occurred about one month after distribution of The Police Chief, and the actual assessment center was scheduled for one month after that. The resume review occurred immediately after the deadline for resume submission. Each successful candidate was

contacted by telephone, advised of his standing and the ensuing steps, and asked if he still intended to participate. If the response was affirmative, any areas of concern noted in the resume review were explored. For example, the reasons for possibly leaving a higher paying position in favor of the Savannah position at lower pay would be discussed. In the month before the center, one candidate withdrew for personal reasons, and a substitute was agreed upon and invited. One hour before the center was to begin, one of the 12 candidates withdrew because of a sudden political controversy in his jurisdiction. The center thus began with 11 candidates, four of whom were from within the agency.

The eleven candidates were divided into one group of six (with two "insiders" of the same rank), and a group of five (with the two other "insiders" of lower rank than in the first group). When the first group was assembled, the environment in the meeting room was tense. The "insiders" were wary of the "outsiders," and vice versa, all of which was compounded by apprehension about the process itself. As the administrator began the orientation (well before the first exercise), one of the inside candidates made an obvious effort to dominate the group, but was countered by an outsider candidate who stated that his absence from his jurisdiction was, in effect, at great personal sacrifice. Assessors noted this interaction, even though it was not technically part of an exercise.

As the assessment center progressed, the candidates in both groups developed the kind of mutual tolerance and respect that grows out of facing and working out problems together, especially in stressful conditions. The assessors reported no observed contamination of the members of the second group by those in the first, although the opportunity existed (the centers were sequentially scheduled). Following the final assessor conference and the compilation of scores from the two groups, the consultant submitted the results to the manager, listing candidates in rank order as determined by their total dimension scores. A narrative summary accompanied each candidate's list of strong and weak dimensions, citing examples of observed behavior that would help in evaluating scores. In this situation, with participants from all over the United States, no feedback interview was possible. The manager, however, used the center results in his subsequent interviews with the competitors from the department.

The results of the center strongly favored two closely ranked candidates, both of whom were from outside agencies. The manager's follow up inquiries and interviews, the latter performed after the center but before the outside candidates returned home, led to the manager's final decision.

The assessment centers in Kansas City, Rochester, and Savannah, each of which was conducted for a different rank, share the favorable reactions of both competitors and those using the results to formulate their selection decisions. Future reviewers will judge the accuracy of those decisions, of course. Initial indications, however, suggest that the centers

matched individual skills and abilities with the important elements of the positions involved. How best to reduce a large number of competitors to the number that can be accommodated in the assessment center is still a problem. The difficulties of attaining a fair, objective, and economical process here clearly indicates that as many candidates as possible should compete in all phases of the selection process.

IV

ASSESSMENT CENTER OPERATIONS

At the Third International Congress on the Assessment Center Method (1975), held in Quebec City, an attempt was made to formulate some "standards and ethical consideration" for assessment center practitioners. The term can be misused; its advocates wanted to avoid the damage to the concept (as well as to people) that would likely occur without some agreement and mutual understanding. A few of the standards seem better suited to the private sector, but most can be used by police agencies at all levels. Besides defining what constitutes an assessment center, the standards include:

- Organizational support for the center through adequate planning, data collection, and professional overview.
- Sufficient assessor training.
- Informed consent on the part of the participants.
- Knowledge about the interpretation of center results on the part of the decision maker.
- A specific, documented validation procedure.

Previous chapters have discussed some of the concerns that are implicit in the summary above, but certain others merit elaboration here.

Assessor Training

Assessor selection and training is probably the most important single element in the method. In Rochester and Kansas City, the assessors were drawn from ranks one level above the position being considered. In Savannah, the assessment staff included well-regarded professionals, although none at the time were chiefs of police. In every case in these examples, the assessors were chosen because they knew the job intimately, and could consistently meet the objectivity required of them. They had some appreciation for the meaning of the police role. While it has been claimed that a well trained assessor can function regardless of personal background or orientation, as a practical matter, credibility must be established. This point can be lost if care is not taken in selecting assessors; Rochester's efforts to curtail potential favoritism paid off.

Interviews with assessors in these cities, after the centers had been performed, highlighted the need for direct practice, the need to apply their new skills experimentally before starting the actual centers.

For example, one department viewed a video taped training session about group discussions. It was a good orientation, but an appreciation for the group dynamics was lost. "It was too impersonal," was the typical comment. In another agency, a few assessors had to be added to the staff after the centers had started; it was expected that sitting in as an observer would be sufficient. This on-the-job training did not work out well, because the new assessors understandably tended to conform to the styles of the assessors they watched, and thus deferred to their more experienced cohorts. Assessor training is most effective if assessors are involved in the job analysis work; but in most cases their normal job responsibilities preclude that. Even intensive assessor training takes at least three days. Insufficient training time was implied when one assessor remarked that "We knew what to look for, but not where to find it."

This mild lament may disappear if, at the start of assessor training, the assessors-to-be help in the actual work of combining task statements and finding acceptable terminology for the behaviors involved. For example, say the task analysis of the position reveals the need for considerable contact with the public and with peers, occasionally under antagonistic circumstances. The ability required is perceiving and reacting appropriately to the needs of others, and objectivity in assessing the impact of self on others. This forms the definition of the behavior, the dimension, that is called "sensitivity."

That name, with its definition, subsequently appears on the report forms used by assessors to report their observations of each exercise. If the relative importance of "sensitivity" in comparision to other dimensions has not been determined, the assessors can mutually agree on its weight for scoring purposes at this point. The training goal here is simply to impart comprehension of the meaning and derivation of the dimensions and, once that is accomplished, to show how the dimensions are revealed in various exercises.

The number of dimensions evaluated depends on the job analysis, but it may make sense to reassign the measure of some to another part of the whole selection process. The 22 dimensions used in Savannah received proper attention only because the assessors were experienced. A dozen or less is realistic for new assessors, provided that other relevant dimensions can be measured elsewhere. For comparison, Chart II shows a list of the dimensions used in each of the three exemplary centers.

In the next training step acquaint assessors with the exercises used. A casual review of their descriptions is not sufficient. Just as development of dimensions by assessors themselves effectively establishes a working knowledge of them, so does involvement of assessors in exercise writing and/or selection.

CHART II

Dimension Comparison

Dimension	Kansas City (Sgt.)	Rochester (Inv.)	Savannah (chief)
Impact	Ability to create a good first impression, to command attention and respect, show an air of confidence and to achieve personal recognition.	Same	Equivalent, adds changes over time
Decisiveness	Readiness to make decisions or to render judgment.	Same	"Decision Making" adds within time frame
Judgment	Ability to reach logical conclusions based on evidence at hand, and to decide when "evidence at hand" is sufficient or more is needed.	Same	Equivalent
Leadership	Effectiveness in bringing a group to accomplish a task and in getting ideas accepted.	Not used	Equivalent, adds a receptivity to ideas
Listening Skill	Ability to pick out important information in oral communication.	Same	Included in oral communications
Oral Communication Skill	Effectiveness of expression in individual as well as group situations.	Add gestures and nonverbal communication	Equivalent: see above
Motivation	Active vs. passive attitude toward career advancement; plans actively for growth.	Not used (see Initiative)	Equivalent
Planning and Organization	Effectiveness in planning and organizing own activities and those of a group.	Same	Adds timetable coordination and use of resources
Problem Analysis	Effectiveness in seeking out pertinent data and determining the source of a problem.	Same	Adds identification of other courses of action
Initiative	Not used	Actively influencing events rather than passively accepting; self starting	See: Motivation
Sensitivity	Not used	Skill in perceiving and reacting sensitively to the needs of others. Objectivity in perceiving impact of self on others.	Equivalent
Stress Tolerance	Not used	Stability of performance under pressure and opposition.	Equivalent
Written Communication Skill	Not used	Ability to express ideas clearly in writing in good grammatical form.	Equivalent; adds brevity

(Other dimensions used in Savannah included integrity, delegating, creativity, emotional maturity, persuasiveness, administrative courage, flexibility, breadth of knowledge, follow up, and sense of mission.)

When an assessor misunderstands exercises, a pattern of superficial observations emerges. The assessor becomes engrossed in who succeeds in a group exercise, for example, rather than the how and why of the participant's success. In an in-basket exercise of several items, the assessor who is not intimately familiar with each item and its relationships with other items, as well as the behaviors signalled by the candidate's responses, tends to rely on the extent of agreement between the candidate's decisions and the assessor's own preferences. Again, the best way to learn about the intricacies of an in-basket exercise is to develop it, but lacking that, instructor critique and explanation along the way are necessary.

A personal discussion between the candidate and an assessor about the candidate's career accomplishments and goals requires practice of even (or especially) an experienced investigator acting as an assessor. The interview focuses on certain dimensions best explored in a one-on-one situation. Each assessor must have adequate time to review biographical data in advance of the interview itself to avoid duplicating factual information already on hand. Advance preparation also helps in the analysis of specific career incidents in terms of the desired dimensions.

With insightful knowledge about the exercises and the dimensions, training should conclude with the direct practice mentioned previously. Questions about report forms and scheduling should be resolved by this point. At least one practice scoring conference should be included, although it can be shortened (in comparison to the real one) without adversely affecting assessor competence. If the center is for career development purposes, assessors should practice the art of interviewing, perhaps with a veteran assessor acting as the candidate.

The fledgling assessor likely will have difficulty concentrating on specific behavioral clues, rather than measuring against some perceived stereotype. With a little practice, however, both formal and informal conversation among assessors soon involves a dimension frame of reference, indicating that assessors have learned the terminology. The use of a consultant to help establish the first training procedure is highly recommended.

Exercise Development

The dimensions having been identified, the agency may elect to develop its own exercises, purchase them, or both. The exercises simply stimulate display of the dimensions; but they will fall short if they do not realistically reflect the job aspired to. The participants hardly can enter into them purposefully if they seem frivolous. They best reflect the contemporary police environment, couched in police language, and contain challenging police problems. At the same time, no competitor subgroup should have an advantage because of familiarity with the kind of problems outlined in the exercise. Candidates usually regard the career interview the least job related, probably because of the direct, deliberate concentration on behavior-revealing experiences, without much emphasis on the police environment per se.

Because all exercises draw out samples of job-related behavior, it is convenient to categorize them by the number of candidates participating, as follows:

- Group (preferably no more than six candidates per group)
- Team (two or three candidates)
- Individual (acting alone)

Group exercises usually include two forms--both without predetermined leadership--designated as assigned role and unassigned role. In the former, the administrator provides certain general information for attacking some given problem common to all roles. Since each participant has a distinctive role, he receives special information also; the assessors note how each candidate makes use of his special information in his interactions with other group members. The latter group exercise, the unassigned role, offers identical information to all participants who jointly work out the stated problem. In both groups, the administrator neither assigns leadership responsibility nor details the specific procedures for arriving at a resolution to the problem. The assessors watch closely to see how, for example, leadership emerges, how it is controlled, maintained, and for what purposes during the exercise. Of course, the assessors evaluate other germane dimensions as they appear.

A variation of a group exercise requires the participants to form competitive sup-groups for part of the exercise, each developing its own strategies for problem resolution. In this situation the assessors must sharply observe not only the sub-group interactions that may take place very quickly, but also the contributions of the individuals evaluated--an especially challenging task. Hence a group should not exceed six candidates in size or a candidate-to-assessor ratio of two to one.

Team exercises call for independent groups of two or three candidates to act together in working out a problem. In a typical situation the candidates may have a set of minimal facts about a police incident. They must piece together in a coherent way what actually took place. A resource person (other than an assessor with evaluation responsibilities for this exercise) responds to the team's questions factually but succinctly. The candidates must decide upon their interrogatory strategy within a time limitation. The resource person may offer some information at a critical point that directly conflicts with the conclusions the team appears to be drawing or has drawn; the assessors observe how the team's members absorb this turn of events. Dimensions such as decision-making, flexibility, communication skills, and planning and organizing are likely revealed in this situation. Rochester included this kind of exercise for the investigator's position, but only one person at a time was evaluated.

Individual exercises usually bring together one assessor and one candidate in review of a completed paperwork problem (in-basket) or of the candidate's career development as reflected in a resume or special interview form. This one-on-one environment offers an opportunity to explore elusive dimensions not adequately revealed in the other exercises. Also, in the single assessor/candidate situation, questions may be answered, schedules or procedures explained, and perhaps the candidate allowed to expound on matters important to him or her.

A variation of the individual exercise used to measure oral skill is to supply factual information to each candidate, and have him deliver his remarks extemporaneously before the assembled assessors. The exercise does not produce many dimension observations beyond oral communication skills and the ability to organize factual material, but these may be quite important to the position under consideration.

One of the advantages of the assessment center methods is its flexible exercise format. The police environment includes many situations which, when simulated in an exercise, will produce valid behavioral observations. However, the exercise creator cannot find ways for the center to measure all of the important skills and abilities required in job performance; the exercise situation must be believable; and finally, the exercises must not deliberately place an individual in a position of ridicule or embarrassment.

Because no one exercise will adequately produce all of the measurable dimensions, the exercise must be developed and combined to elicit as many of the dimensions as possible, and as often as possible. Draw up a dimension grid, listing the dimensions in one column and the exercises arrayed on the other axis. A mark at each intersection denotes an opportunity (usually) to observe the dimension. This guides the assessors as they prepare for each exercise, and assures satisfactory coverage of the dimensions.

The exercises simulate, and thus require verification of their validity. They should be standardized. Accordingly, they should be field tested before used in an actual competitive situation. The candidates must not be in a position to compromise test security. Also, group and team exercises should not exceed an hour's duration; the in-basket or other written exercises should be completed within three hours. The number and variety of report forms required of assessors can prove vexing. For each exercise, assessors must make quick observations of behavior, and in a form amenable to recall later. Other forms summarize observations for the assessor compiling feedback information, and still others to record the participants' rankings of their peers and assessors' rankings of participants in group exercises. There is no ready solution other than experimentation with various formats, but each listing of the dimension, other than on an administrator's tally sheet, should include the definition of the dimension as well.

The Administrator

Each assessment center requires an overall coordinator. While this person may contribute observations during the assessors' conference, mainly he or she should orient the participants, control and disseminate exercise materials, instruct participants in exercise procedures, consolidate assessors' report forms, debrief participants, and serve as a moderator at the final assessors' conference. A veteran assessor may fill both assessor and administrator roles, but it is likely that both roles will suffer.

The orientation is particularly important. The candidates should be introduced to their assessors, with a brief description of the assessors' credentials. The candidates may know one another; if not, wait until after the first group exercise to introduce them to each other, to avoid the problem of deference to a higher ranking participant. Most assessment centers randomly assign neutral identification badges (colors are typical) to participants, to enhance equality among participants and to foster a new environment with no special status indicators. Requiring casual clothes is useful, both to reduce the impact of the stressful setting and to enhance equality.

The administrator describes the exercise schedule and perhaps some of the background about assessment centers, emphasizing that participants be themselves (i.e., not try to fake behavior, because it is highly unlikely that acting will be successful for the duration of the center). At this point, perhaps take pictures of the participants (a Polaroid type camera is ideal) for use in the final assessors' conference. The tone of the orientation should be as nonthreatening as possible, despite pictures being taken. As the center proceeds, the administrator begins and ends exercises, tallies ranking forms, makes sure candidates and assessors are in the right places at the right times, fends off the media on occasion, and, in general, sees to the center's smooth functioning.

Scheduling

The candidates in a competitive environment naturally feel stress. However, the schedule first must meet the assessors' needs. This means adequate time after each exercise at least to note quickly the more important behavioral examples, to review an in-basket exercise thoroughly before the follow-up interview, and to break for meals.

Mere words cannot convey the feelings and irrational behavior of assessors who have conscientiously observed, probed, and written reports during an improperly scheduled center. It is tempting to schedule an assessment center for a long day (extending into the evening), but several such centers in a row soon demoralize assessors.

The schedule should list the candidates each assessor is to evaluate for each exercise. To minimize subjectivity, assessors must observe different candidates in each exercise. Also, in the situations where groups include candidates well known to each other, the administrator should try to avoid assigning closely associated candidates to the same group discussions or team exercises. Similarly, avoid assigning superiors and direct subordinates to the same groups or teams. But even without scheduling flexibility, the center will still produce adequate opportunities for behavior observation. The assessors, in this case, should know about prior relationships, and thus watch for any proclivity to support a friend or superior, for example, in preference to personally enhancing behavior.

Tangible evidence of deference to another candidate at the expense of personal standing may exist in peer rankings. Immediately after each group exercise, the administrator asks each candidate to rank, on a supplied form and in private, his peers and himself in order of each participant's overall effectiveness during the exercise. (There may be other rankings requested also, on criteria such as contributions to the resolution of the problem, the degree of sensitivity to other participants, and so forth.) Once collated, these rankings may reveal unrealistic evaluations by candidates.

Chart III shows the results of rankings by candidates and assessors (who also rank participants in the same manner) for a group of six after an hour's group discussion. The final rankings exclude the self-ranking by a candidate and add the remaining numbers in each row. The lowest total number equates to top ranking, and so forth. In this example, the assessors generally agreed with the candidates' rankings; within the candidates' group, there is general agreement with the exception of Black's ranking of Green and Blue. Black seems to promote Blue's interests despite contrary evaluations by his other peers, while assigning himself an unusually low place. Such a ranking would be checked by the assessor responsible for observing Black during the exercise, and would be further discussed--with the evaluation confirmed or contradicted--during the assessors' conference (as Black's dimension scores are reported for the other exercises).

Occasionally, in order to balance schedules and assignments, and round off a group, it is tempting to add an off-duty assessor or other noncompetitor posing as a candidate. Never do this; the ethical ramifications, as well as the unmeasurable biases that may be introduced, must not be ignored for the sake of expediency.

Facilities

Neutral identification for participants creates a common environment, and the same applies to the actual assessment center site. If possible, a hotel or motel away from headquarters is the most desirable

CHART III

PEER RANKING EXAMPLE

LEADERLESS GROUP DISCUSSION #2 January 2

*Indicates self-ranking

Candidates' Rankings

Candidate	Ranked by: Red	White	Blue	Green	Black	Brown
Red	2*	2	3	1	2	2
White	3	3*	4	4	5	4
Blue	6	6	2*	5	1	5
Green	1	1	1	2*	4	1
Black	5	4	6	6	6*	6
Brown	4	5	5	3	3	3*

Assessors' Rankings

	Assessor: Whitney	Hickox	McDonald
Red	2	2	2
White	5	3	4
Blue	6	6	6
Green	1	1	1
Black	4	5	5
Brown	3	4	3

Final Rankings By:

	Assessors	Candidates
1	Green	Green
2	Red	Red
3	Brown	Brown) tie
4	White	White) tie
5	Black	Blue
6	Blue	Black

facility. Have a sufficiently large meeting room for each group exercise and the in-basket exercise, and private rooms for the individual career and in-basket follow-up interviews. In any event, security for exercise materials and results must be provided. Privacy is desirable, of course, but may become critical when competitors do not want their participation known.

The Assessor Conference

Second only to assessor training in importance, the assessor scoring conference brings together the individual behavioral observations of the assessors to form overall evaluation for each candidate. Research shows that this evaluation, the end of the process, is more accurate than any particular component. As the typical conference starts, with only the administrator and assessors present, the assessors have completed all their exercise reports and have reviewed the ranking forms for each group exercise.

The assessors discuss the candidates one at a time, with each assessor responsible for observing an exercise, making his observations and the scores for each dimension he observed known to the group. The frame of reference is always a dimension. Other assessors may request elaboration or challenge his findings, but in any event they note the scores reported by the assessor in relation to their own, if they differ. The usual scoring scale, from one to five, denotes quantities of the dimension ranging from very little to a great amount. After all the exercise observations have been reported, the assessors independently determine a final reporting score for each dimension: They review the behavioral observations noted in the exercises that produced a measure of the dimension and the reporting assessors' scores. The administrator then, in random order, calls on each assessor for his final score for each dimension and tallies the results.

The give-and-take during this evaluation is important, because it helps assessors recognize other indicators of behavior, and may actually compensate for temporary lapses. In the early stages of assessor development, assessors may differ by a spread of two or three points on a particular dimension. The administrator must encourage resolution of this disparity by probing the basis for the scores -- not simply averaging them.

The total of all the dimension scores (weighted if appropriate) provides a rank ordering, if in a competitive situation. The dimension scores also reveal each candidate's strengths and weaknesses and, when coupled with observed examples of behavior, serve as the basis for feedback, or as evaluative aids in further selection options. The Appendix illustrates three of the reporting forms used during the scoring conference.

Assessment Center Resources

This monograph describes various considerations in the use of assessment center methodology. Obviously much effort must go into establishing and operating a center. Consider the steps taken in most cases:

- Determining the need for a center and the end products
- Civil service involvement and/or clearance
- Conformity with departmental regulations and/or procedures
- Allocation of financial support
- Selection of consultants, if needed
- Job/task analysis
- Skills and abilities determination
- Dimension development and weighting
- Procedures for fairness/validity/reliability evaluations
- Exercise selection
- Exercise development and standardization
- Determining weight of assessment center in total evaluation
- Informing candidates
- Arranging facilities
- Selecting and training assessors
- Determining methods of reducing the candidate field
- Administering the center
- Compiling and reporting results to appointing authority
- Candidate feedback/debriefing
- Verification of validity

In summary, the operation of an assessment center in a police agency demands a substantial commitment. The experience of those agencies that have used the method suggests that such an effort is worthwhile.

The complexity of the process justifies the use of consultants in at least the initial phase. One way to evaluate the potential for local use is to visit an assessment center in operation, not only to gain greater understanding of the method, but also to ascertain the need for consultants.

Several private consultants and at least one professional association offer assessment center services. Exercise packages are also available. Over the long term, however, it may be wise to develop internal capability, including the creation of exercises; this allows better control over the security of center exercises and reporting materials, and permits tailor-made exercises using local setting--all of which address the credibility problem.

Finally, this monograph cannot explain every aspect of assessment center operation. It should help in the process of evaluating assessment center potential for local use, and as a guide for overall development, whether undertaken with or without a consultant. No mere publication, however detailed and exhaustive, can develop and refine the interpretive skills that assessors must bring to the evaluation of candidates. The greatest appeal of the assessment center method is that those charged with evaluative responsibilities can rely upon an understandable, valid methodology.

Appendix

Because the dimensions related to each position differ, it is impractical to devise various assessment center reporting forms for all situations. The general format and application of the most important scoring forms should be noted, however. Shown here are:

- The assessor's report form
- The assessor's tally sheet
- The administrator's tally sheet

ASSESSOR'S REPORT FORM

Candidate Green Date June 12

Exercise In-Basket

Dimension	Score
1. PLANNING AND ORGANIZING. The ability to implement a decision through development of a workable program, so as to achieve desired results.	2

(Observations) Green failed to plan for the needed patrol coverage of the threatened demonstration (Item #11, and also failed to organize the In-Basket material at the start. Made little use of org. chart, routed memos and directives improperly.

Assessors use this form to record their dimension scores for each candidate for whom they are responsible on each exercise. They also note specific examples of the behavior that led to the scoring decision, in this case a "2".

The assessor completes the form as soon after the exercise as possible, from brief notes made on a special exercise form that facilitates notetaking. The assessor charged with conducting the counseling interview after the center may use observations from the form in discussions of the candidate's strengths and weaknesses.

ASSESSOR'S TALLY SHEET

Candidate Green Date June 13

Your Score 2

Dimension

1. PLANNING AND ORGANIZING. The ability to implement a decision through development of a workable program, so as to achieve desired results.

Group discussion #1	2	-	In-Basket	-	2
Group discussion #2	3	2	Interview	-	-
				A	Y
Oral presentation	3	-			
	A	Y			

A = Reporting assessor's score
Y = Your score

Each assessor uses this to record his evaluations as well as those of other assessors, for each dimension, as the assessor's conference proceeds. In this case, the dimension of planning and organizing" for candidate Green is being discussed. This assessor was responsible for evaluating Green's in-basket exercise, awarding a score of "2" on the basis of the quantity of the dimension shown. On group discussion #1, the responsible assessor reported a "2"; this assessor agreed with that score, based on what the other assessor described and what the first assessor had seen and heard, although the first assessor was responsible for other candidates during that exercise. For group discussion #2, the responsible assessor seemed too generous in awarding a "3"; hence, the first assessor noted a "2" on the form. The other scores are reported and evaluated in like manner.

After all the assessors' reports have been discussed, each assessor independently inspects all the reported scores (because "planning and organizing" is not measured on the personal interview, no socre is entered) and decides upon a final mark. This assessor decides that the dimension score of "2" will stand for the final tally.

Candidate _____ Date _____

Dimension	Assessor A	B	C	s/t	Wt	Total
1. Planning and organizing	2	2	2	6	1.3	
2. Sensitivity	3	4	3	10	1.5	
3. Leadership	3	3	3	9	2.0	
4. Oral communication skills	2	3	2	7	1.5	
5. Judgment	3	3	4	10	1.3	

After the assessors have discussed all the exercise dimension scores for a candidate and independently decided on their final dimension scores, the administrator calls for a score from each assessor, dimension by dimension, and records them in the manner shown above. The weights determined earlier in the job analysis pahse then apply, and the administrator calculates the score for rank ordering purposes.

As suggested in the tally above, occasionally assessors report scores separated by three or more points on the usual scale of one to five. For example, assessor B originally reported a "4" for oral communication skills, while A and C each reported a "2". The administrator then reviews the scoring rationale with each assessor, and reaches an agreement about the fairest score. In this case, B elected to lower the score to a "3," but A and C might have raised their scores to either "3" or "4" in light of B's explanation.

Such disparity does not imply a lack of understanding or inconsistent standards; rather, it usually indicates that performance on one exercise unduly impressed an assessor.

Candidates occasionally display an unusual amount of a particular dimension on one exercise, but the same dimension is hardly present in others. The abrupt change may come of a deliberate attempt to compensate for perceived losses in other exercises, deference to a peer, and so on. By pooling assessor observations over the duration of the center, however, quite likely an accurate evaluation will emerge during the scoring conference.