1978

Analysis of the use of video tape in the assessment center method

David Carrington Purdy

ANALYSIS OF THE USE OF VIDEO TAPE IN THE ASSESSMENT CENTER METHOD

BY

DAVID CARRINGTON PURDY

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
OF THE UNIVERSITY OF RICHMOND
IN CANDIDACY
FOR THE DEGREE OF
MASTER OF ARTS
IN PSYCHOLOGY

JUNE 1978

ANALYSIS OF THE USE OF VIDEO TAPE IN THE ASSESSMENT CENTER METHOD

BY

DAVID CARRINGTON PURDY

Approved:

Robert J. Filer, Thesis Director

Jean N. Dickinson

Warren P. Hopkins

L. James Tromater, Chairman,
Dept. of Psychology

Table of Contents

Preface

The author would like to express his sincere gratitude to all those whose contributions made this study possible. In particular, many thanks are due Dr. Robert J. Filer who provided the initiative for the project, the exercise materials, and the technical expertise in assessment center methodology. His constant encouragement is greatly appreciated. The assistance of Dr. Warren P. Hopkins in reading the manuscript and suggesting improvements was quite beneficial. Many thanks also to Ms. Jean N. Dickinson who served on the examining committee.

Special thanks are due Dr. John Hamilton who helped to coordinate the project at Virginia Polytechnic Institute and State University. The author is very grateful to the Hamilton family for their warm hospitality during the many weekend stays in Blacksburg. Considerable recognition is due to the assessors: Ms. Jamie Carlisle, Mr. Lee Elliott, Mr. Thomas Garwood, Mr. Edward Regan, Mr. Edward Trautman, and Mr. James Utterback who spent much of their Winter and Spring quarters at the assessment centers. Their assistance is greatly appreciated.

List of Tables

Abstract

This study examined the feasibility of employing video-taping procedures in the operation of assessment centers. Six trained assessors observed 23 participants in four separate assessment centers. While one group of three assessors observed the assessment proceedings in person, the other group of three saw only the video-tape playbacks of the performances. A high degree of reliability was found between the ratings given by the two teams of assessors. The Pearson Product Moment Correlation computed on the total scores of all 23 participants yielded a reliability of $r = .86$. The Spearman rank-order correlation coefficient was $r = .85$. Video-taping procedures did indeed convey the essential behavioral information necessary to evaluate an individual's management potential. Positive and negative factors of video taping are discussed in the paper.

Analysis of the Use of Video Tape in the Assessment Center Method

One of the most rapidly proliferating techniques in the management development field is the assessment center method of identifying and developing management potential. An assessment center is not a place, but, rather, a program in which participants proceed through a number of different exercises which are designed to simulate management activities. These evaluation procedures include in-basket exercises, management games, leaderless group discussions, and interviews. These job-related simulations are designed to elicit behaviors which are relevant to dimensions previously determined to be crucial to job success. At the conclusion of the center, a report is prepared which presents the results of each candidate's performance. Organizations using the assessment center method can be found in every major industry including government and in almost all the industrialized nations of the world. This relatively new concept in assessment is being used at all levels within organizations, from line supervisors to top management.

The modern-day assessment center has its roots in the multiple assessment procedures utilized by German military psychologists prior to and during World War II. The importance which the Germans placed on assessing individuals within the context of real-life group situations has survived as an essential characteristic of today's assessment programs. In 1942, the British adopted many of the techniques used by the Germans and formed the British War Officer Selection Boards (WOSB). One year later, the American Office of Strategic Services (OSS) began to use the method for selecting intelligence agents. In the private sector, American Telephone and Telegraph was the first to use the method and has assessed more than 100,000 candidates since 1956. Today, AT&T operates fifty centers throughout the country, processing about 10,000 candidates per year. The assessment center method has also been widely applied in governmental agencies. Federal courts have accepted the assessment center method as a valid and nondiscriminatory means of evaluating performance potential. In Richmond, Virginia, the courts have approved the adoption of the assessment center method for determining promotions within the police and fire bureaus.

There is a substantial body of evidence to document the validity of the assessment center technique. A survey of published research studies by Cohen, Moses and Byham (1973) indicates that the median correlation between assessment center ratings and supervisors' ratings of potential for further promotions to higher management following the first promotion is .63. Some of the best evidence for the validity of the technique comes from the AT&T Management Progress Study (Bray & Grant, 1966), which evaluated 422 individuals by the assessment center method. Each person was given a rating on 25 dimensions as well as an overall rating that represented a consensus judgment regarding the candidate's future management potential. In the five to seven year span of time since assessment, 82% of the college men who had reached middle management positions had been correctly identified by the assessment center. In addition, the assessment center correctly identified 94% of those individuals who never advanced past first-level management positions.

While the assessment center has gained general acceptance as a uniquely useful and successful method, one of its major drawbacks is the high cost factor. There is wide variation in estimates of the costs involved in running an assessment center, but a reasonable figure seems to be approximately $500 per candidate (Filer & Filer, 1977). This figure would include time lost by assessors and candidates from their jobs, as well as the lodging and food for twelve participants, six assessors, and one administrator. Travel expenses to the actual assessment center location can also be quite high. In 1969, a division of IBM reported a cost of $5000 per twelve assessees (which included motel expenses); staff salaries were not included (Byham, 1969). Due to organizational differences and diverse methods of expense allocation, it is difficult to generalize about costs associated with an assessment center. A number of factors are involved in determining the overall costs: the number of assessees, the number of consultants employed, the actual place of assessment (whether on company premises or at a luxury hotel), the materials used, and the length of time that candidates and assessors are off the job.

One possible solution to this high cost factor is the employment of video tapes to record the candidates' performance in the assessment center exercises. This would cut food, lodging, and travel expenses significantly.

At the conclusion of the assessment center exercises (one or two days), the administrator would simply transport the tapes to trained assessors who would then view the performances via closed-circuit television. Besides cutting costs dramatically, this approach would also allow for the replaying of certain segments of the tapes for clarification purposes. In addition, the tapes could be used for feedback purposes. This would provide valuable information to the participants concerning their performance during the assessment center. Used in this manner, video taping could serve as a powerful training and development tool.

The purpose of this study is to explore the reliability of using video tape in the operation of assessment centers. Research in other areas has shown the use of video tape to be effective and reliable. Miller (1975) demonstrated that jurors who viewed live vs. video-taped presentations responded comparably (with the exception that the jurors in the video-tape group retained more trial-related information). Similar findings were reported by Chesley (1974) who concluded that the use of video tape in court and in legal proceedings was "a method of the future" (p. 162). He found that video tapes of witnesses' statements and of medical operations were useful in courtroom presentations. Research by Waters (1975) suggests strong similarity of data from video-taped interviews and from face-to-face interview procedures. He found that the subjects responded to the video-tape recorded interviewer as if the interviewer were present in person. Moore and Lee (1974) compared interview trait ratings made by managers who viewed video-taped interview playbacks with ratings made by experienced interviews in 34 live interviews. Mean ratings were similar in both the live and video tape groups. They maintained that video tape facilitated interviewer reliability because of the replay feature, and concluded that the high level of agreement between face-to-face and video-taped ratings indicated that perceptual distortion was not a strong factor. Hall (1967) has gone so far as to state that "the television camera is more sensitive than the human eye and can provide a better image than an observer can perceive when viewing a live interview" (p. 23).

In the field of drug research, findings by Newmark, Dinoff, and Raft (1974) indicate that the standardized video-tape interview shows promise of being a highly reliable and clinically valid instrument for the purpose of assessing behavioral effects in psychotropic drug studies.

Spencer, Corcoran, Allen, Chinsky, and Veit (1974) investigated the use of the video tape for observing the interactions on a ward for retarded children. They found that video taping can give an accurate representation of ward activities as viewed by trained observers in the real-life milieu. They compared ratings made by observers who were present in person with those made by observers who only viewed the video tapes. Concordance ranged from 54% for the context dimension to 85% for the initiator dimension. When the video-tape observers were given a second viewing of the tapes, the concordances rose to 70% and 89%. The average cross-modal agreement was 79%. Their findings indicated that the video-tape procedure allows greater flexibility with regard to observer scheduling and inadvertant observer absences. In addition, providing multiple exposure to taped interactions improved efficiency in training new raters, and the researchers suggested that this would increase reliability estimates of trained observers.

Results of a study carried out in a therapy situation (Eisler, Hersen, & Agras, 1973) indicate that the reliability of video observations is as high as for live observations of the same behavior. Their results clearly indicated that video-tape observation of nonverbal interaction for looking and smiling behaviors of married couples is highly reliable and equal to reliabilities obtained by observing the interactions live. They emphasized a distinct advantage of video tape: interactions can be replayed numerous times to focus on additional behaviors which were not perceived live. In addition, the use of video tape facilitates precision in defining and measuring behaviors during subsequent replays.

While there have been no studies concerning the reliability of video-tape procedures in the assessment center method, a great deal of research has been conducted on inter-rater reliability. Some of the earliest work concerning the reliability of assessment center-type procedures was conducted in Great Britain by Vernon and Parry (1949). They reported an overall agreement of .80 for ratings on 125 recruits by two separate War Office Selection Boards.

More recently, there have been a number of studies showing sufficiently high inter-rater reliabilities in assessment center evaluations to justify continued use of the method. In the Management Progress Study (Bray

& Grant, 1966) inter-rater reliabilities were .75 for both ratings and rankings when two psychologists observed a leaderless group discussion. The reliabilities on ratings and rankings based on the observation of a manufacturing problem were .60 and .69.

Greenwood and McNamara's (1967) research at IBM resulted in high inter-rater reliabilities for ratings and rankings of three different pairs of alternating observers. The median of the 432 reliability coefficients computed for ratings was .74 and .76 for rankings. They also computed inter-rater reliabilities separately for different pair combinations of psychologists and line managers. They found no significant differences in degree of agreement between certain types of assessor pairs.

The SOHIO Program (Thomson, 1970) yielded similar findings. Inter-rater reliabilities for ratings by two psychologists on 13 dimensions ranged from .73 to .93, $\bar{r}$ = .85. In addition, there was high agreement between psychologists' and managers' ratings.

The preceding studies were concerned with comparing ratings given by single assessors and also by pairs of assessors. A more recent study which compares ratings made by two separate teams of three assessors was conducted by Michigan Bell Telephone Company (Moses, 1973) to study the relationship between two multiple assessment programs. A selection of 85 nonmanagement employees first attended a one-day assessment center program. At least one month later, the same participants were assessed by a more extensive two-day program. Independent groups of assessors were used, and the second group was not given any information concerning the individuals' performance in the first center. The study yielded a substantial correlation between overall performance in the two centers for the total sample (.73) as well as for each of the subgroups (men, .77; women, .70; blacks, .68; whites, .73). There were no significant differences found between the reliabilities obtained for these four subgroups. The results also yielded a correlation of .56 between the two competitive group exercises.

Richards and Jaffee (1972) have shown that a substantial increase in reliabilities is achieved by training observers. When untrained observers assessed individuals on a human relations skill dimension, the mean inter-rater reliability was .46 versus .78 for trained observers.

Some might maintain that the high reliabilities reported in the research on assessment centers are artificial because of assessor training methods and the assessors' discussion of rating differences (Huck, 1973). They might suggest that reliabilities should be computed on independent judgments. However, every performance appraisal has as its goals common standards and accurate interpretation of performance data. The assessment center method puts great emphasis on training assessors, which appears to be well-justified in light of the above findings.

This study is designed to examine the feasibility of employing video-taping procedures in the operation of assessment centers. It is hypothesized that reliability coefficients will be similar to those computed in studies which compare two separate groups of assessors which view the same assessment proceedings in person. These have typically ranged from approximately .60 to .85.

Method

Subjects. Six graduate students (ages 22-33) in applied behavioral science who had received intensive training in behavioral observation served as assessors. Five of these were male and one was female. Their participation in the study was voluntary, and several of them received practicum credit. Twenty-three students at Virginia Polytechnic Institute and State University volunteered to be candidates for assessment. They were recruited primarily from business classes. As an incentive, they were promised a full report on their performance in the assessment center. In addition, they were given the opportunity to view the video tapes of their performance and to receive verbal feedback from one of the assessors. Eleven of the participants were undergraduates (juniors and seniors) and twelve were graduate students. All but seven assessees were enrolled in the business school. Sixteen participants were male while seven were female. Ages ranged from 20 to 33. Four assessment centers were held with six individuals evaluated each time. During the first center, one volunteer failed to appear, reducing the number of assessees to five.

Materials. Four exercises were used in the assessment center. A two-hour managerial in-basket called "Top Flight" consisted of 22 different memos, letters, and notes which the participant, in the role of a general manager, had to take action on. A half-hour interview simulation required

the participant to hold an interview with a problem subordinate (acted out by a role-player) and to convince him to change several negative behavior patterns on the job. Finally, all assessees participated together in two leaderless group discussion exercises. "Automobile Design," an exercise with unassigned roles, required the six participants to reach a consensus agreement within one hour regarding the design of a new automobile. The second group exercise, "Superior Products," involved assigned roles, and each participant was given information on one individual whom they were to defend as a candidate for promotion. The group was given one hour to reach a conclusion. Each participant was provided with their own copy of the exercises.

Assessors were provided with numerous instructional materials during the training program which were organized into an "Observer's Manual." This was used as a resource manual during the actual assessment centers. The administration of each center required the use of three large rooms and six smaller cubicles. Two Sony monochromatic video cameras were used along with a Sony cassette recorder and a Hitachi reel-to-reel recorder. In addition, two monitors were used, one with a 21-inch screen and the other with a 12-inch screen. Each assessment center required the use of eight one-hour tapes which were re-used during each subsequent center. The group exercises required a round table large enough for six people to sit around one side such that each individual could be "seen" by the camera.

Procedure. Prior to the first assessment center, the six assessors underwent a three-day training session in which they developed skills in observing specific behaviors and categorizing these into relevant behavioral dimensions. For practice, they each took the in-basket exercise and later evaluated one which had been completed by a role-player. They studied all of the exercises and then had the opportunity to observe each one as it was acted out by role-players. During all of these practice exercises, they focused on one individual whom they actually evaluated. The assessors also practiced holding the assessors' discussion in which evaluation was made of the individuals' performance along critical dimensions.

At the end of the discussion, each assessor wrote a final report which summarized the strengths and weaknesses of the participant, and included an agreed-upon consensus as to the applicant's potential for further development. The experimenter concluded the training session with feedback

concerning the assessors' behavioral observations, discussion sessions, and final reports.

Following training, two pilot assessment centers were run in which improvements were made in the video-taping process and in the administration of the exercises. Assessors gained experience in the operation of the assessment center and in making behavioral observations.

After the two pilot sessions, four separate assessment centers were conducted, with four different groups of six candidates (only five appeared for the first center). The same six assessors served at each center: three observed the participants in person, while the other three observed the exercises via video-tape recordings. Each assessor had the opportunity to serve twice in both the live condition and the video-tape condition. To insure heterogeneity of groups, assessors were rotated between each group of three as an added precaution. In other words, the same three assessors never served together more than once. Table 1 shows the assessor assignment schedule and rotation formula.

Table 1

Assessor Assignment Schedule and Rotation Formula

Assessment Center 1	Assessment Center 2	Assessment Center 3	Assessment Center 4
Live Group:			
Assessor A	C	A	D
Assessor B	D	E	B
Assessor C	E	F	F
Video Group:			
Assessor D	A	D	A
Assessor E	B	B	E
Assessor F	F	C	C

Each assessment center was conducted in the following manner. After the six candidates had arrived at the center, the administrator gave a welcoming talk and overview of the assessment center method. Participants signed consent forms (Appendix A) and then began working on the two-hour in-basket exercise. Following this, they were given 15 minutes to prepare for a 30-minute interview with a problem employee, role-played by one of the video-equipment technicians. Two interviews were conducted simultaneously in separate rooms and recorded on video tape. There was one

assessor in each room during the interviews making observations. At the conclusion of this exercise, participants were free to leave for the evening. Each assessor spent the remainder of the evening evaluating two in-baskets and completing the exercise report forms. They also prepared to conduct an in-basket interview for the next day.

The following morning, all six participants were given background information relating to the "Automobile Design" exercise. After spending 15 minutes preparing their arguments and studying the information, the participants had one hour to reach agreement on the problem at hand. The three assessors sat in the room and each observed two participants. The video camera was in plain view although the monitor was kept in the adjoining room to avoid distraction. At the conclusion of this group discussion, the participants were interviewed by an assessor in regard to the actions they took in the in-basket exercise. Each interview lasted thirty minutes and was also video taped. Following this, candidates were given 15 minutes to prepare for the final exercise, "Superior Products." All six participants again assembled around the large round table and had one hour to agree upon a solution to the problem. As in the other group discussion, all three assessors sat in the room and made their behavioral observations. The video camera was positioned in the same place as before, but the candidates were seated at different places around the table. Each assessor had the opportunity to observe all six participants during the course of the assessment center, but had primary responsibility for the two whose in-basket exercises he/she evaluated. Feedback was also provided to these same two individuals by the assessor.

In each of the exercises, the video camera was pre-focused and the tripod was locked in place. Both the audio and video levels were pre-checked and permanently set. Selection of the image was a wide-angle shot, permitting a view of participants' entire bodies when they were in a sitting position. During the course of the exercises, the angle of the camera was not changed, and no lens changes were made.

At the conclusion of all exercises, the assessees were debriefed and arrangements were made for their feedback sessions. Those expressing an interest were allowed to return in four days to view the video playbacks of their performance in the assessment center. After the participants had

departed and the assessors had completed all of their exercise report forms, the assessors' discussion was begun.

The objective of the final assessment discussion was to obtain an overall objective evaluation of the individual's potential to perform along the ten critical dimensions. The discussion was held by three assessors evaluating six participants. One participant was discussed at a time with each assessor reading to his/her fellow assessors his report of the participant's performance during the exercises. While an assessor presented his exercise report, the other assessors took notes. Following the reading of behaviors listed under a particular dimension, each assessor gave a rating for that dimension on a one to five scale. (The dimension evaluation guides and the operational definitions of each dimension are included in Appendix B.) Discussion then followed in which assessors arrived at a consensus rating of that particular dimension. One dimension was discussed at a time until all dimensions were completed. In a similar fashion, the three assessors arrived at an overall consensus rating for the participant regarding their total performance in the center. After all dimensions were completed, the assessor responsible for the final report collected all exercise reports and other information applicable to the participant's performance in the center. These final reports summarized the assessor's discussion and indicated the strengths and weaknesses of the participant as well as an agreed-upon consensus as to the individual's potential for managerial functioning. This same procedure was followed for each of the six participants in the center.

After the "live" group of assessors had completed their work, the "video-tape" group of assessors was ready to begin the same procedure, although they had only the video tapes of the exercises and the written products (e.g. the in-baskets) to evaluate. The three assessors in this latter group had no personal contact with the participants and made all of their observations based strictly on the video playbacks of the exercises. In addition, they were provided with the actual in-baskets which had been completed by the participants.

As they observed the video playbacks, they were free to stop the tape at any point in order to record more thorough observations. They could also backtrack and review segments of the tape for clarification. Seating

charts showed the position and name of each participant so that there was no confusion as to each assessee's identity. The same steps as described for the "live" group were followed in this group with the exception that no final reports were written by the assessors in the "video-tape" group. The video group typically spent their first day observing the exercises via tape and completing their report forms. On the second day, they participated in the assessors' discussion. The experimenter requested that there be no communication whatsoever between the two groups of assessors concerning the participants.

This same procedure and sequence of events was followed for all four assessment centers.

Results

The Pearson Product Moment Correlation computed on the total scores of all 23 participants yielded a reliability coefficient of $r = +.86$, $p < .001$. This total score is a summation of the ratings for all ten dimensions. This high correlation of total scores between the two independent groups of assessors indicates a strong degree of reliability in using video tape. Assessors also gave a global consensus rating to each participant, and the reliability coefficient of these 23 ratings was $r = +.80$, $p < .001$. The Pearson correlation coefficient for all 230 ratings (23 participants X 10 dimensions) was $r = +.64$, $p < .001$.

Table 2

Pearson Reliability Coefficients, Level of Significance, Means and Standard Deviations for Ratings by Live and Video Assessor Groups

	N	r	Level of Significance	Mean		Standard Deviation	
				Live	Video	Live	Video
Overall Consensus	23	.8027	0.001	5.7391	5.4348	1.0539	1.3760
Total scores (sum of all ten dimensions)	23	.8572	0.001	57.3478	54.6087	9.9937	12.8726
All participants on all dimensions	230	.6367	0.001	5.7348	5.4609	1.4400	1.6628

Table 3

Pearson Reliability Coefficients, Level of Significance, Means and Standard Deviations for Ratings on Ten Dimensions by Live and Video Assessor Groups

Dimension	N	r	Level of Significance	Mean		Standard Deviation	
				Live	Video	Live	Video
Leadership	23	.6275	0.001	5.4348	4.9565	1.5323	1.5805
Planning and Organizing	23	.4966	0.008	5.2609	5.3913	1.4528	1.9941
Management Control	23	.6012	0.001	5.1304	4.6957	1.8167	1.8692
Analytical Skill	23	.3973	0.030	6.1739	5.9130	1.2668	1.4744
Sensitivity	23	.3976	0.030	5.6957	5.5652	1.1051	1.2368
Decisiveness	23	.6320	0.001	6.1304	5.7391	1.2175	1.2869
Stress Tolerance	23	.7025	0.001	5.8261	5.8261	1.3702	1.3366
Activity Level	23	.7803	0.001	6.0435	5.5217	1.5219	1.8554
Oral Communication	23	.7806	0.001	5.9565	5.6957	1.0215	1.6358
Written Communication	23	.7586	0.001	5.6957	5.3043	1.7434	2.0323

Table 3 summarizes the reliability coefficients computed for each dimension. These ranged from $\underline{r} = .78$ for oral communications skill to $\underline{r} = .40$ for analytical skill. The highest degree of consistency between the two assessor groups was found on the following dimensions: oral communications skill, activity level, written communications skill, stress tolerance, leadership, and management control. These coefficients were all significant at the 0.001 level. There was less agreement on sensitivity, analytical skill, and planning and organizing.

The data was further analyzed using the Spearman rank correlation method. The 23 participants were ranked according to their total scores on all ten dimensions. Where ties existed, mean ranks were assigned to sets of tied individuals. This analysis yielded a Spearman rank-order correlation coefficient of $\underline{r} = .85$.

The Kendall tau coefficient is a somewhat different approach to the problem of agreement between two rankings. While Spearman's method treats the ranks as though they were scores and then finds a correlation coefficient, the computation of tau depends on the number of inversions in order for pairs of individuals in the two rankings. There is a relationship between the two conceptions, but they are by no means identical. In Spearman's method, the process of squaring differences between rank values in $\underline{r}_S$ places different weight on particular inversions in order. Kendall's tau puts equal weight on all inversions by a simple frequency count (Hays, 1973). The Kendall tau correlation between the overall ranking of all 23 participants is tau = +.72. Again, a very high degree of agreement is found between the live and video groups of assessors.

Discussion

This study was undertaken to analyze the feasibility of using video tape in the assessment center method. It was hypothesized that the video-tape method would be reliable. The significant correlations derived in comparing the ratings given by the "live" group of assessors with those given by the "video" group indicate that consistent standards were employed in both methods and that video presentations do indeed convey the essential behavioral information necessary to evaluate an individual's management potential. Not only were the correlations of all 230 ratings high, but the rank order correlation of the 23 participants was particularly high.

These results compare quite favorably with Moses' (1973) reliability study in which 85 individuals were assessed by two different teams of assessors. There was no video taping involved, and they came up with the correlation of .73 between the two independent assessments.

It was interesting to discover that there were differences among the various dimensions in the degree of agreement between the live and video groups. Particularly high correlation coefficients were found for the ratings given on oral communications skill, written communications skill, and activity level. It would be expected that agreement should be high on written communications effectiveness since the in-basket was the only exercise in which this dimension could be observed. Both the live and the video groups of assessors evaluated the in-basket material in the same manner and the television medium was not involved. The correlation coefficient for oral communications was even higher, which indicates that both verbal and nonverbal cues were communicated effectively via video tape. There was also strong agreement between the groups on the dimension called activity level.

The lowest correlations were found on the dimensions analytical skill and sensitivity. A possible explanation for this follows. The exercises chosen for this study were selected partially on the basis of how well they lent themselves to video taping. In a typical assessment center, a specific one- to two-hour analytical exercise is included which gives a measure of an individual's logical thinking ability. This sort of exercise was not included in the present study since it is an individual written exercise which would not lend itself to video taping. The only opportunity the assessors actually had to rate this dimension was in evaluating the participants' arguments in the group discussions, the notes they prepared prior to the exercises, and their actions taken in the in-basket. When compared to a specifically designed analytical exercise, these other exercises are somewhat weak for assessing an individual's analytical ability.

There seemed to be some misunderstanding among the assessors concerning the definition of the dimension called sensitivity, which is defined as "skilled in listening to others and reacting sensitively to their needs in a tactful and understanding manner." This definition caused problems because it lacks specificity and leaves too much open for

interpretation. Evidently, the assessors were not using consistent standards in determing their ratings on this trait. In future training sessions, particular attention should be directed towards specifying in behavioral terms what constitutes high and low ratings in sensitivity. A more precise definition should also be provided to the assessors so that consistent standards can be maintained from one center to the next. Improved dimension evaluation guidelines with greater behavioral specificity would be very useful.

It is important to note that those dimensions which are often weighted double or triple in promotional assessment centers had relatively high correlation coefficients. Leadership, management control, and decisiveness all had correlation coefficients above .60, all significant at the 0.001 level.

While the results of this study do indicate that video tape can be reliably used in the operation of assessment centers, there are a number of considerations which must be weighed concerning its feasibility under a given set of circumstances. There are two major apparent disadvantages: technical problems and assessor problems.

During the course of this study, a number of technical difficulties were encountered in the video-tape system. With so many different pieces of equipment, failure in one unit often caused the breakdown of the entire system. This often resulted in delays in the schedule while the problems were being diagnosed and corrected. In order for video tape to be used effectively in assessment centers, well-trained technicians are a necessity. They must be able to trouble-shoot and make necessary repairs quickly. It is also recommended that an inventory of extra cords, microphones, and other recording equipment be kept on hand as back-up in case of failures. Effective lighting is also a crucial factor in making good quality recordings.

The assessors also cited a number of drawbacks to the video-tape method. They found it tedious to observe the one-hour group exercises on the television screen, and explained that they missed actually being in the room as the exercises unfolded. They also missed having personal contact with the participants in the assessment center when serving in the video group. The feeling of personal involvement was absent. They also

found that it took a longer than normal period of time to associate participants' voices with their faces. It was found that large-screen monitors were essential for effective observation of the group exercises. Small screens reduced the size of faces to the point that nonverbal behaviors were difficult to detect.

There was one particular problem in using just one camera in the group exercises. The two individuals at either end of the semi-circle were at a disadvantage in that the camera captured only a profile of their faces. This posed somewhat of a problem to the assessors. By using two cameras at different angles and two monitors for playback, this problem could be avoided.

The assessors did note advantages to the video-taping method. They found it a great help to be able to rewind the tape and re-view segments of the tape when they were unsure of some particular behavior. By stopping the tape at any point, they were able to make more complete observations without having to fear missing something else. This feature of video taping was particularly useful in the group exercises when each assessor was observing two participants. Additional behaviors could be rated subsequently on replay. The behavioral observations appeared to be richer from the assessors when serving in the video group. Assessors also found it helpful to increase the volume level during playback which facilitated the understanding of oral communications, particularly for participants who spoke softly. While assessors in the live group may have had difficulty understanding the quiet-spoken participants, the assessors in the video group could compensate for this by increasing the volume.

The results of this study have many implications for the future spread of the assessment center method. Reductions in the cost factor through the use of video tape make the assessment center more feasible for smaller companies. Consulting firms can offer a lower price figure by taping the center proceedings on the client company's premises and then replaying the tapes back at their own home office. Professional assessors could then view the tapes on a more flexible schedule. Travel and lodging costs could be greatly reduced since assessors could be local residents. This type of arrangement avoids the high cost of transporting, lodging, and boarding assessors. Regional assessment centers could also become more viable through the use of video tape.

Initial start-up costs would be fairly high with the purchase of good quality video-tape equipment. In the long run, however, the savings realized would quickly compensate for the initial outlay. Further testing of the video tape approach in business and industrial settings is the next step in establishing its practicality. The potential for creative uses of video tape in the assessment center method is great. Where cost and logistics factors have made the use of the method prohibitive to smaller corporations in the past, new possibilities are now opened for wider applications of the assessment center method.

APPENDIX A

Informed Consent Form

Informed Consent Form

David Purdy and/or Dr. John Hamilton have explained my participation in this experiment. I am aware of the following points and my participation is voluntary.

1. I will be participating in four assessment center exercises in which my behavior will be observed by three trained assessors. The proceedings will be recorded on video tape for later viewing by another group of assessors. These tapes will remain confidential.

2. I will be furnished a summary report outlining my strengths, weaknesses, and developmental needs as indicated by my performance in the assessment center. These reports will be written by the assessors and approved by Dr. Robert Filer or Dr. John Hamilton.

3. All information from this experiment will become the property of the Department of Psychology and will be accessable only to those involved in this project. Although results of this experiment may be made public, my identity will be disguised to insure anonymity.

4. I can terminate my participation in the experiment at any time.

_____

Signature of Participant

_____

Witness

_____

Date

APPENDIX B

Dimension Evaluation Guide

DIMENSION EVALUATION GUIDE

Leadership

Effectiveness in bringing a group to accomplish a task and in getting ideas accepted. Commands attention through respect and personal accomplishment.

Rating	Definition	Behavioral Examples
5	Superior--high leadership qualities demonstrated throughout the exercise	1. Took charge of group and was clearly identified by other participants as the leader. 2. All suggestions and recommendations were accepted.
4	Above average--leadership qualities greater than normally expected from people in this position	1. Took an active part in directing the activities of the group and was respected by other members. 2. Most suggestions and recommendations were accepted.
3	Average--leadership qualities typical of people in this position	1. Occasionally led the group discussion. 2. Some suggestions and recommendations were accepted.
2	Below average--leadership qualities below what would normally be expected in this position	1. Rarely held the group's attention. Attempts at leadership were usually rejected. 2. Few suggestions and recommendations were accepted.
1	Poor--leadership qualities far below what would normally be expected in this position	1. Made no attempt to lead the group, or all leadership efforts were rejected. 2. Made no suggestions, or none were accepted.
X	Dimension not observable in this exercise	

DIMENSION EVALUATION GUIDE

Planning and Organizing

Effective in planning and organizing own activities and those of a group.

Rating	Definition	Behavioral Examples
5	Superior--very strong planning and organizing ability demonstrated throughout the exercise	1. Read through all in-basket items before taking any action and divided them into several piles on the basis of priority. 2. Notes on interview simulation anticipated responses of interviewee and outlined alternative courses of action.
4	Above average--planning and organizing qualities greater than normally expected from people in this position	1. Read through in-basket items and acted on most important ones first. 2. Plan for the interview simulation indicated clear order of topics to be covered.
3	Average--planning and organizing qualities typical of people in this position	1. Skimmed in-basket although he failed to see a few priorities. 2. Prepared a rough sketch of topics to be covered in the interview simulation.
2	Below average--planning and organizing qualities below what would normally be expected in this position	1. Skimmed in-basket exercise, but failed to note correct priorities in many cases. 2. Plan of interview simulation was disorganized and did not cover major topics adequately.
1	Poor--planning and organizing qualities far below what would normally be expected in this position.	1. Made no attempt to set priorities for items in the in-basket exercise. 2. No plan of interview simulation was apparent.
X	Dimension not observable in this exercise	

DIMENSION EVALUATION GUIDE

Management Control

Establishes methods of control and monitors results and activities of his subordinates to insure that objectives and directives are met.

Rating	Definition	Behavioral Examples
5	Superior--strong management control demonstrated throughout the exercise	1. In the in-basket exercise set up follow-up meetings to monitor performance of subordinates on all items delegated. 2. Established several ways to follow-up the activities of subordinate in the interview simulation.
4	Above average--management control qualities greater than normally expected from people in this position	1. In the in-basket exercise established follow-up procedures on most items. 2. Established definite follow-up procedures with subordinate in the interview simulation.
3	Average--management control qualities typical of people in this position	1. Established follow-up procedures on approximately half the items delegated to subordinates in the in-basket. 2. Suggested the possibility of a follow-up with his subordinate in the interview simulation.
2	Below average--management control qualities below what would normally be expected in this position	1. Established follow-up procedures on very few items in the in-basket. 2. Suggested that subordinate "get back to him later" but did not establish a definite time.
1	Poor--management control qualities far below what would normally be expected in this position	1. Made no attempt to establish follow-up procedures on any items in the in-basket. 2. Made no suggestion of follow-up to subordinate in the interview simulation.
X	Dimension not observable in this exercise	

Analytical Skill

Conclusions reached are logical and based on effective analysis through seeking pertinent data and determining the source of the problem.

Rating	Definition	Behavioral Examples
5	Superior--strong analytical qualities demonstrated throughout the exercise	1. In the in-basket exercise saw all relationships among items and took appropriate action. 2. All claims assignments were made appropriately and the choice of a promotion was based on good logical analysis.
4	Above average--analytical qualities greater than normally expected from people in this position	1. Saw the interrelationships of most items in the in-basket and took appropriate action. 2. Made only one error in assigning claims to representatives and selection of promotion based on sound logic.
3	Average--analytical qualities typical of people in this position	1. Missed a few key interrelationships items in the in-basket and failed to take appropriate action on these. 2. Made several errors in assigning claims and some faulty logic in making promotional decision.
2	Below average--analytical qualities below what would normally be expected in this position	1. Failed to see many of the interrelationships among in-basket items and failed to take appropriate action on these 2. Made numerous errors in assigning claims and showed faulty logic in the choice of an individual to be promoted.
1	Poor--analytical qualities far below what would normally be expected in this position	1. Saw none of the interrelationships of items in the in-basket. 2. Made incorrect assignments of all claims and selection of promotional decision was totally devoid of logic.
X	Dimension not observable in this exercise	

Decisiveness

Ready to make decisions or to render judgment. Tenacious in staying with an important problem or decision until the matter is settled or reduced in priority.

Rating	Definition	Behavioral Examples
5	Superior--high degree of decisiveness demonstrated throughout the exercise	1. In the group exercise stuck by all his recommendations even when questioned by the majority of the group members. 2. In the interview simulation established clear direction for his subordinates and was not influenced by the subordinate's arguments.
4	Above average--decisiveness greater than normally expected from people in this position	1. Stuck by his recommendations in the group exercise unless overwhelming opposition was encountered. 2. In the interview simulation was firm with his subordinate but yielded on some minor points.
3	Average--decisiveness typical of people in this position	1. Stuck by some recommendations in the group exercise but gave in on others. 2. In the interview simulation agreed with the role player on about half the points.
2	Below average--decisiveness qualities below what would normally be expected in this position	1. Rarely stuck by his arguments in the group exercise. 2. In the interview simulation agreed with the role player a majority of the time and did not state specific corrective action.
1	Poor--decisiveness qualities far less than normally expected in this position	1. Quickly abandoned his arguments in the face of opposition in the group exercise. 2. Allowed the subordinate to completely dominate the interview simulation.
X	Dimension not observable in this exercise	

DIMENSION EVALUATION GUIDE

Sensitivity

Skill in listening to others and reacting sensitively to their needs in a tactful and understanding manner.

Rating	Definition	Behavioral Examples
5	Superior--high sensitivity demonstrated throughout the exercise	1. In the interview simulation took care to insure that all decisions were truly acceptable to his subordinate. 2. In the group exercise tactfully acknowledged that all arguments of other group members had some validity.
4	Above average--sensitivity greater than normally expected from people in this position	1. Indicated understanding and empathy on most problems presented by the subordinate in the interview simulation. 2. In the in-basket exercise attempted to strike balance between job requirements and subordinate feelings.
3	Average--sensitivity typical of people in this position	1. Was generally tactful but belittled arguments of some group members in the group exercise. 2. Listened sympathetically to the subordinate in the interview simulation but insisted on his own point of view.
2	Below average--sensitivity below what would normally be expected in this position	1. In the in-basket exercise would not allow a subordinate to participate in marriage counseling. 2. Frequently belittled arguments of others in the group exercise.
1	Poor--sensitivity far below what would normally be expected in this position	1. In the interview simulation completely disregarded the employee's explanations of the problems. 2. In the group exercise refused to acknowledge the validity of any other points of view.
X	Dimension not observable in this exercise	

Activity Level

Is a self-starter and maintains a high activity level.

Rating	Definition	Behavioral Examples
5	Superior--very high activity level demonstrated throughout the exercise	1. In the in-basket exercise completed all items and wrote extensive memos and notes on most problems. 2. Made numerous suggestions in the group exercises and spoke more than any other participant.
4	Above average--activity level greater than normally expected from people in this position	1. Completed all items in the in-basket although explanations on some items were fairly brief. 2. Made recommendations on all problems in the group exercises.
3	Average--activity level typical of people in this position	1. Acted on all but one or two items in the in-basket. 2. Completed all assignments in the Winnco Claims exercise, but most explanations were quite brief.
2	Below average--activity level below what would normally be expected in this position	1. Failed to complete numerous items in the in-basket and appeared to work quite slowly. 2. Made few suggestions and participated little in the group exercises.
1	Poor--activity level far below what would normally be expected in this position	1. Completed less than half of the in-basket items and worked quite slowly. 2. Did not participate at all in the group discussions.
X	Dimension not observable in this exercise	

DIMENSION EVALUATION GUIDE

Stress Tolerance

Performance is stable under pressure and opposition.

Rating	Definition	Behavioral Examples
5	Superior--high level of stress tolerance demonstrated throughout the exercise	1. In the group exercises did not appear to be at all disturbed despite vigorous criticism and opposition. 2. Maintained stable composure and relaxed manner in interview simulation despite angry outbursts by subordinate.
4	Above average--stress tolerance greater than normally expected from people in this position	1. Was mildly upset but maintained control well under vigorous questioning in the group exercise. 2. Became somewhat flushed when questioned in the in-basket interview about inappropriate decision but quickly regained his composure.
3	Average--stress tolerance qualities typical of people in this position	1. Became irritated at times in the group discussion when challenged but did not withdraw. 2. Was somewhat nervous when making presentations in the confrontation exercise.
2	Below average--stress tolerance qualities below what would normally be expected in this position	1. In group exercise was noticeably nervous and somewhat disorganized when giving presentations. 2. In the interview simulation jumped from point to point and did not answer the role player's questions directly.
1	Poor--stress tolerance qualities far below what would normally be expected in this position	1. In the in-basket interview became so upset that he was unable to answer questions. 2. Exploded angrily in the group session under mild questioning by peers.
X	Dimension not observable in this exercise	

Oral Communications Skill

Able to express ideas clearly and in a persuasive manner in oral communications.

Rating	Definition	Behavioral Examples
5	Superior--strong oral communications skill demonstrated throughout the exercise	1. Responses to questions in the in-basket interview were well constructed and thoroughly addressed the point in question. 2. Spoke in a clear, loud, and fluent manner and was easily understood by all participants in the group exercise.
4	Above average--oral communications skills greater than normally expected from people in this position	1. Answers to questions in the in-basket interview were generally adequate although some minor points had to be re-asked. 2. Presentation in the Winnco Claims exercise required only minor follow-up questioning.
3	Average--oral communications skill typical of people in this position	1. Was necessary to ask for further explanation to several of his responses in the in-basket interview. 2. Other group members occasionally asked him to speak up in the group exercise.
2	Below average--oral communications skills below what would normally be expected in this position	1. Voice inflection and accent distracted from presentation. 2. Used poor grammar and mixed syntax.
1	Poor--oral communication skill far below what would normally be expected in this position	1. Failed to make self understood even after repeated rephrasing in the in-basket interview. 2. Unable to explain actions in the Winnco Claims exercise in an understandable manner.
X	Dimension not observable in this exercise	

DIMENSION EVALUATION GUIDE

Written Communication Skill

Able to express ideas clearly and in a persuasive manner in written communication.

Rating	Definition	Behavioral Examples
5	Superior--strong written communication skill demonstrated throughout the exercise	1. Written communications in the in-basket and Winnco Claims exercises were thorough and concise and no follow-up questions were necessary.
4	Above average--written communication skill greater than normally expected from people in this position	1. Written explanations in the in-basket and claims exercises were generally clear although minor questioning on some points was necessary.
3	Average--written communications skills typical of people in this position	1. Written communications in the in-basket and claims required additional explanation of some major points in some items.
2	Below average--written communication skill below what would normally be expected in this position	1. Written communications demonstrated poor grammar and organization and required considerable questioning for understanding.
1	Poor--written communication skill far below what would normally be expected in this position	1. It was impossible to tell what the participant intended from an analysis of his written communications in the in-basket and Winnco Claims exercises.
X	Dimension not observable in this exercise	

References

Bray, D. W. & Grant, D. C. The assessment center in the measurement of potential for business management. Psychological Monographs, 1966, 80 (17, Whole No. 625).

Byham, W. C. The uses of assessment centers. A summary of material presented at the Spring, 1969 meeting of the Executive Study Conference.

Chesley, S. M. The use of video tape: The expert witness. Journal of the American Institute of Hypnosis, 1974, 15, 161-162.

Cohen, B. M., Moses, J. L., & Byham, W. C. The validity of assessment centers: A literature review. Journal of Industrial and Organizational Psychology, Summer, 1973.

Eisler, R. M., Hersen, M., & Agras, W. S. Video tape: A method for the controlled observation of nonverbal interpersonal behavior. Behavior Therapy, 1973, 4, 420-425.

Filer, R. J. & Filer, R. K. Assessment Centers: Development and Use. Psychological Consultants, Inc., 1977, 42-43.

Greenwood, J. M. & McNamara, W. J. Inter-rater reliability in situational tests. Journal of Applied Psychology, 1967, 31, 101-106.

Hall, W. O. Analysis of video tape in the perceptual evaluation of empathic communication. Unpublished manuscript, 1967.

Hays, W. L. Statistics for the Social Sciences. New York; Holt, Rinehart and Winston, 1973.

Huck, J. R. Assessment centers: A review of external and internal validities. Personnel Psychology, 1973, 26, 191-212.

Miller, G. R. Juror's responses to videotaped trial materials: Some recent findings. Personality and Social Psychology Bulletin. 1975, 1, 561-569.

Moore, L. F. & Lee, A. J. Comparability of interview, group, and individual interview ratings. Journal of Applied Psychology, 1974, 59, 163-167.

Moses, J. L. The development of an assessment center for the early identification of supervisory potential. Personnel Psychology, 1973, 26, 569-580.

Newmark, C. S., Dinoff, M., & Raft, D. The standardized video tape interview as an objective dependent variable in psychotropic drug research. Journal of Nervous and Mental Disease, 1974, 158, 18-24.

Richards, S. A. & Jaffee, C. L. Blacks supervising whites: A study of inter-racial difficulties in working together in a simulated organization. Journal of Applied Psychology, 1972, 56, 234-240.

Spencer, F. W., Corcoran, L. A., Allen, C. R., Chinsky, A. M., & Veit, T. S. Reliability and reactivity of the video tape technique on a ward for retarded children. Journal of Community Psychology, 1974, 2, 71-74.

Thomson, H. A. Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.

Vernon, P. E. & Parry, J. B. Personnel Selection in the British Forces. London: University of London Press, 1949.

Waters, T. J. Further comparison of video tape and face-to-face interviewing. Perceptual and Motor Skills, 1975, 41, 743-746.