INSTRUMENTS BIAS IN ASSESSMENT CENTERS

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

H. Michael Cunningham
Denton, Texas

August, 1984

EB

Cunningham, Howard Michael, Instrument Bias in Assessment Centers. Doctor of Philosophy (Clinical Psychology), August, 1984, 70 pp., 5 tables references, 43 titles.

The purpose of this study was to examine the effects of behavioral checklist critical item content on subsequent global, Likert-type ratings. It was hypothesized that assessment center participants rated with positive critical items would receive higher scores on subsequent global ratings than would participants rated with negative critical items. Additionally, it was hypothesized that volunteers would receive better ratings than nonvolunteers. Finally, it was hypothesized that behavioral ratings would show less susceptability to halo effect than global ratings. Subjects were eight male and 14 female volunteers and nonvolunteers, recruited from graduate business classes. They were randomly assigned to be rated with positively or negatively stated behavioral checklists. New assessment center technology which reduces the influence of participant "likeability" and augments the method's content validity was used. Dependent measures were global ratings of management level, management potential, and the difference between assessor's and participants' rating of management potential. While the first hypothesis was not supported, the other two were confirmed. No differences were found between participants rated with positive versus negative

critical items. The lack of significance for this factor was probably due to the primacy effect, because all raters were trained on the positively stated checklist before introducing the negatively stated checklist. Volunteers were rated as significantly better on measures of management level and management potential. The average intercorrelation between dimensions on the behavioral checklist was smaller than that for global ratings, indicating a reduction in error variance attributable to halo effect. The findings demonstrate the behavioral checklist's superiority over global ratings in achieving nonbiased ratings. Implications of these findings as well as suggestions for further research were discussed.

TABLE OF CONTENTS

Table of Contents--continued Page

INSTRUMENT BIAS IN ASSESSMENT CENTERS

The Management Assessment Center has emerged as the preferred method of management selection in settings ranging from large private sector corporations to small, independent companies. As the Equal Employment Opportunity Commission demands for nondiscriminatory selection techniques increase, personnel officers have turned to this technology to insure valid and legally acceptable selection decisions. The assessment center is a cost effective, content valid selection device, withstanding the challenge of courtroom debates (Firefighters Institutes for Racial Equality vs. City of St. Louis, 1980; Friend et al. vs. Leidinger et al.,[City of Richmond], 1977).

Definition and Description

William Byham (1975) provides a definition of the assessment center. He states that

> The assessment center is a formal procedure incorporating group and individual exercises for the identification of dimensions of managerial or sales success which have been identified as important for a particular position or level of management (pp. 497-498).

The typical assessment center is a two to three day procedure wherein participants are involved in myraid of exercises that simulate actual managerial tasks. These exercises are usually

separated by periods during which psychometric inventories are administered, although in some centers which use managers rather than psychologists as raters no such inventories are given. Managers are typically evaluated on dimensions such as energy, resistance to stress, oral and written communication skills, and the functions of organizing and planning (Huck & Bray, 1976).

The Task Force on Assessment Center Standards (1980) provides criteria considered necessary for an adequate assessment center. This task force recommends the following elements be included in the center:

1) Multiple assessment techniques, at least one of which must be a simulation exercise. A simulation exercise is one in which stimuli representative of real life situations are presented. The exercise is intended to elicit those behaviors related to performance in the real life situation. Representative simulation exercises used in the assessment center include group exercises, in-basket exercises, and interview simulations.

2) Multiple assessors. Assessors must be thoroughly trained observers familar with targeted dimensions of behavior and their operational definitions. Training and familiarization with these dimensions and the rating format must be completed prior to the assessment center process itself.

3) Pooled ratings. The overall assessment center rating must reflect the ratings from all assessors across all assessment center techniques. This pooling is done after the actual assessment center. All conclusions and recommendations concerning a participant should be based on the pooled ratings.

4) Prior analysis of relevant job behaviors. A job analysis must be done prior to designing the assessment center to insure that dimensions, attributes, characteristics, qualities, skills, abilities, and knowledge to be evaluated and presented in simulation exercises are in fact relevant to the job for which the participants are being considered (p. 37).

Thus, an assessment center is not a place, but a technique. Paper and pencil tests may be utilized, possibly even a full battery of aptitude, ability, and personality tests. Interviews and simulation exercises are also used. However, these techniques used separately do not make an assessment center. The essential ingredient is the combination of these selection methods. The assessment center is a mixture of simulation exercises and other techniques that have been determined to be relevant to the targeted position.

Court Rulings on Assessment Center Validity

As stated by the Task Force on Assessment Center Standards (1980), simulation exercises are the backbone of the assessment center technique. The court system has consistently upheld

employment decisions which are based on simulation exercises. In one case (Friend et al. vs. Leidinger et al. [City of Richmond], 1977), the judge found paper-and-pencil selection techniques to be biased racially, but when used in conjunction with the powerful simulation exercises, determined that the overall selection technique was nondiscriminatory. Since the assessment center selected 21.4% of the Blacks and 21.3% of the Whites applying for the position of fire lieutenant, no adverse impact was demonstrated. Furthermore, the judge ruled that the assessment center was valid and job related for this position. Thus, even if equal percentages of Blacks and Whites were not found acceptable, the selection procedure would still be deemed appropriate. Using this technology, the city could demonstrate that "adverse impact is a necessary but unintentional result of job-related requirements" (Friend et al. vs. Leidinger et al. [City of Richmond], 1977, p. 361). The court dealt with this case as if the City of Richmond were a private employer, setting an important precedent for future discrimination suits in both the public and private sectors.

In another case (Firefighters Institutes for Racial Equality vs. City of St. Louis, 1980), the assessment center was exonerated in an otherwise biased selection battery. While requiring the City to bring its selection procedures into compliance with government standards, the judge mandated that all immediate promotion decisions be based soley

on the assessment center portion of the battery. The decision held simulation exercises to be content valid for the job at hand.

Benefits of the Assessment Center Method

Certainly, these court decisions laud the content validity of the assessment center. Additionally, factors beyond content validity add to the method's utility. If outside assessment centers are used, as opposed to "in-house" assessment, the assessors typically will not know the participants. Thus, objectivity is increased over supervisor ratings (Byham, 1970). The use of multiple raters increases the reliability of the assessment information. Further, by employing psychologists as well as managers as assessors, the observational skills of the former can be combined with the greater business knowledge of the latter to produce a more well rounded picture of the participant (Byham, 1975; Kraut, 1972).

An additional benefit of the assessment center method is that participants receive clear, detailed feedback regarding their strengths and weaknesses (Byham, 1975), making staff development more efficient (Byham, 1970). However, many managers fear a poor rating, believing that below standard performance will result in a "kiss of death." Byham and Pentecost (1970) suggest that managers who perform poorly leave their companies at an unusually high rate. On the other hand, Huck (1973) was unable to verify this fear.

He agrees with Kraut (1972) that a poor rating is not a "kiss of death" nor does an excellent rating create a "golden boy." A commission on ethical use of the assessment center method (Moses, 1975) recommended that management inform participants as to how the assessment results will be used. This information should eliminate unwanted apprehension which may interfer with work performance.

Validity Studies

In an effort to document predictive validity as well as content validity, Byham has evaluated 22 research validity studies (Slevin, 1972; Byham & Wettengel, 1974). In 15 of the studies acceptable validity was demonstrated. Unfortunately, Byham neglected to report the actual number of subjects used in each study. Assessment center techniques have also been found to have predictive validity for minorities and women (Byham, 1975; Huck & Bray, 1976).

In a review of 18 studies conducted between 1964 and 1972 Huck (1973) found the assessment center method to be an accurate predictor of job potential, job progress, and job performance. Longitudinal studies with thousands of employees have also shown the assessment center process to be more accurate than traditional methods in predicing "hard criteria" such as promotions, performance, and salary increases (Byham, 1970). Similar findings were obtained regardless of whether or not the companies knew about the assessment center results when giving promotions and raises.

Bray compared men whose promotions were based on assessment center results with men who were promoted in the usual manner (Huck, 1973). He found that 67% of the assessed versus 35% of the nonassessed managers were judged "better than satisfactory" on the job. Unfortunately, Bray neglected to report the significance of the difference in these percentages.

Byham (1975) assessed 143 college-educated and 144 non-college-educated managers, using the assessment center method. The results of the assessment were not made available to the company as these men advanced through their careers. The assessment center identified 32% of the college and 75% of the non-college managers who were later rated "successful" and 88% of college and 95% of noncollege managers who were later rated "not successful." These findings suggest that Byham's assessment center is better able to identify failures that successes. Perhaps his raters have been trained to notice behaviors that are predictive of failure rather than success.

Does the assessment center have predictive validity for all levels of managers? Byham (1970) argues that predictors are better for midlevel managers than for entry-level managers although significant predictions can be made for low-level managers (Huck, 1973). In fact, the technique shows more potential for managers advancing to higher levels even if the assessment center is targeted for entry-level positions.

For a procedure to be truly advantageous, it must not only accurately predict success and failure, but also outperform existing methods of selection. Byham and Wettengel (1974) report that the assessment center is better than paper-and-pencil tests and panel interviews. Additionally, all three methods may be combined to boost criterion-related validity (Byham, 1975). Other authors (Huck, 1973; Wollowick & McNamara, 1969) state that the use of simulation exercises can double the variance accounted for, using "success" as a criterion, over paper-and-pencil tests of management ability alone. They also find that over 80% of the variance in the assessment center ratings is independent of scores which are based on an analysis of personnel records.

On the other hand, some research has shown that assessment centers do not improve the quality of manager selection over traditional methods. Some authors (Byham, 1975; Wilson & Tatge, 1973) conclude that assessment centers are cost-efficient only if the candidate either had no job history or if the new position was considerably different from his/her current position. Otherwise, the person's past performance in similar jobs was the more accurate predictor. These authors also contend that predictions based on psychological inventories are more accurate than those based soley on simulation exercises.

Weaknesses in Current Methodology

Even with the documentation of content and predictive validity, a few isolated studies support other methods of management evaluation. Several weaknesses in currently practiced assessment center procedures deserve attention. First, in traditional assessment centers, the participants have ample time to prepare for each simulation exercise. Therefore, the requirement to respond as they would on the job can not be fully evaluated. Secondly, since each dimension is evaluated by global ratings, a consideration of differential performance in each exercise and the assessment of different management styles is lost. Global ratings are also subject to irrelevant variables such as the participant's "likeability" (Bray, 1982; Cunningham, 1982; Lany & Farr, 1980), suggesting that such ratings should be abandoned. These global ratings have been also shown to result in a positive bias when raters were trained to identify positive managerial traits. Additionally, global Likert-type rating scales have been shown to be especially vulnerable to halo effects (Cooper, 1981). Further, the techniques now used do not allow in-depth assessment of the participant's time management skills nor the ability to integrate material from several exercises when making decisions. Certainly these are valuable management skills. Surprisingly, many validity studies fail to report the number of subjects and do not adequately describe the exact assessment procedures. Finally, much assessment

center data is reported as percentages of managers found to be "successful" rather than providing hard statistical findings. These methodological problems severely restrict the value of the assessment center research conducted in the past.

Methological Remedies

One solution for these shortcomings was to construct an integrated series of exercises which simulate a "day at the office" (Smith & Cunningham, 1982). Each exercise draws on information from previous exercises, allowing assessment of participant's ability to integrate material across tasks. Since participants are rated after each exercise, the effects of accumulated stress and fatigue can be evaluated. Stress forces the participant to rely on deeply ingrained coping strategies rather than cognitively mediated responses. Thus, participant's performance in latter exercises is assumed to be indicative of the way they will perform in a stressful managerial position.

A behavioral rating scale was also constructed by Smith and Cunningham (1982) that allows raters to indicate how the participant handles each task. The use of such a scale also improves the ability to provide detailed feedback for evaluative and remedial action (Jacobs & Zedeck, 1980). This rating scale was found to eliminate a portion of the bias related to "likeability" which is inherent in subjective, global ratings (Cunningham, 1982). Finally, accepted statistical

procedures were used to demonstrate significant effects. Smith and Cunningham (1982) report that these procedural changes increase the method's content validity while shortening the time needed to assess each participant, making the assessment center procedure more powerful, cost-efficient, and scientific.

Byham (1970) offers evidence that such an approach is warranted. He stated that although each assessment center is different, "the accumulation of research findings from a variety of types of centers lends considerable credibility to the general validity of the technique" (Byham, 1970, p. 154). The short, one-day centers are about as accurate as the three-day procedure. Moses (1973) found that correlations between one and three-day assessment centers cluster around .68 to .77. Therefore, as long as simulation exercises are incorporated, various methodological schemes should be acceptable as long as certain requirements remain. These findings indicate that even though validity has been shown, we can still improve many aspects of the methodology.

Cognitive Set and Rater Bias

Since Smith and Cunningham's (1982) procedures have been shown to be superior to traditional methodology for controlling some biasing factors, their techniques should be used in future assessment center research. Because of the control over biasing effects, this methodology is especially well suited for studying one important variable that has been overlooked

in the literature, rater bias. Cunningham's (1982) major finding that behavioral checklist rating significantly reduces rater bias which is based on participant "likeability" raises the question of bias inherent in the use of behavior checklists. Since the categories and critical behaviors are identifed in advance, raters may selectively tend to focus on certain behaviors and discount other behaviors. Thus, "cognitive set" may influence the assessors' perceptions of the participants (Eisen, 1979).

The identification of positively stated critical items may have contributed to the positive bias that Cunningham found in the global ratings. In fact, Srull and Wyer (1979, 1980) have shown that as the number of observed behaviors which are consistent with a particular personality trait increases, the assessor's accessibility to that trait as a general classification and organizing theme also increases. Increasing accessibility to a trait causes an increase in the likelihood that the trait will be used to integrate subsequent behavior. Thus, the presentation of a checklist with delineated behaviors will probably create a cognitive set that affects the way future behaviors are observed and classified.

Pioneering work in the area of cognitive sets was done by Asch (1946) and extended by Kelly (1950). By giving subjects the expectation that a target person was warm or cold, these authors were able to influence the subject's perceptions

and subsequent ratings of the target person. Srull and Wyer (1980) further demonstrated that ratings can be influenced positively or negatively by manipulating the trait terms made accessible to raters. These findings have been extended to include clinician's ratings (Huguenard et al., 1970) and global ratings of instructors (Nisbett & Wilson, 1977). A review of the literature indicates that assessment center researchers have overlooked this factor, however.

It seems that there are several mechanisms responsible for this type of rater bias. After forming initial impressions from a small sample of behaviors, raters tend seek confirmatory evidence (Darley & Fazio, 1980). Raters will test their hypotheses about the participant by selectively attending to confirming behavior. They will even slant their own behavior so as to elicit behaviors from the participant that are consistent with the raters initial impression Snyder & Swann, 1978). Additionally, ambiguous behaviors on the part of the target will probably be interpreted as being consistent with this impression (Darley & Fazio, 1980). Kelly (1950) has shown that if favorable rather than negative impressions are formed, subjects will interact with targets more frequently. Confirmatory seeking behaviors on the part of a rater may elicit behavior from the participants that is inconsistent with their normal responses. Thus, questions asked during an interview, subtle or nonverbal cues during exercises, and general attitude toward the participant may vary depending on which trait categories are primed.

Other cognitive factors also influence this process. Confirmatory behaviors will be rememberd as occuring more frequently than nonconfirmatory behaviors (Synder & Swann, 1978). Raters will even "remember" nonpresented material that confirms their impression of the target (Cantor & Mischel, 1977). "Impression-maintenance attributional bias" causes rater to assume that confirming behaviors are a dispositional quality of the target and that disconfirming behaviors are a function of situational forces (Darley & Fazio, 1980). For example, if a rater primed with positive managerial traits sees a participant emerge as a leader in a leaderless group, the bias would be to assume that the participant has strong leadership ability. However, if the rater later saw that participant's leadership usurped by another participant, the bias would be to assume that something in the situation caused a temporary lapse in his ability to take control. Thus, category availability influences the rater's cognitive set which, in turn, effects the rater's future behavior and the ways in which they process subsequent information.

Will distortions causes by observing previously delineated management behaviors result in a distorted global picture of an assessment center participant? Srull and Wyer (1980) conclude that the "priming" the rater with trait categories will influence future judgements on traits originally primed as well as on related traits. Global ratings have

been shown to be based on those qualities deemed central to the person (Mensh & Wishner, 1947). Since centrality or peripherality of traits depends on the situational context, it is expected that management behaviors will come to be viewed as central during a management assessment. As time passes, raters will tend to rely more on the central categories than on specific behaviors when making rating judgements. This information processing bias is consistent with "information reduction mechanisms" observed and documented in related cognitive research (Cantor & Mischel, 1977).

The effect of such cognitive bias is seen not only in the direction (positive or negative) of the subsequent ratings, but also in increased intercorrelations between rating categories (Cooper, 1981; Jacobs & Zedeck, 1980). Reliance on global impressions increases the likelihood that ratings on supposedly unrelated categories will covary. Thus, those assessment centers using behaviorally anchored, global ratings (Huck & Bray, 1976) are especially vulnerable to this biasing effect.

If global ratings are affected by the types of categories made accessible to raters, all previous research in assessment center technology can be questioned. If the raters in the various studies were not trained to look for the same participant behaviors, this cognitive bias will have affected their ratings so that the studies cannot be directly compared. Further, validity and reliability data for an assessment

center using one training method (or type of checklist) may not hold for centers using other methodologies.

Behavioral Checklists and Cognitive Set

Assessment centers using behavioral checklists to rate participants may also be vulnerable to bias resulting from cognitive set. "Primed" categories which result in cognitive bais may be determined by the actual critical items used on behavioral checklists.

The purpose of the assessment will influence the types of critical items used. Selection and promotion decisions may be based on positively stated items. Termination decisions may use negatively stated critical items. Finally, an overall, general assessment center may use a combination of both types of items. Therefore, findings may not be comparable from one assessment center to another (Jacobs & Zedeck, 1980). The lack of available literature on the effects of differing critical items on subsequent ratings leaves assessment center directors who are dissatisfied with global ratings in an untenable position with respect to this issue. Thus, research in this area is sorely needed.

Assessment centers generally should be grouped according to their purpose. The types of critical behaviors to be measured should be identified and all critieria operationally defined. Finally, awareness of the bias inherent in any rating form or training method should help the conscientious assessment center director devise strategies that will lead

to a more fair, accurate estimate of the participant's management potential.

Purpose and Hypotheses

The purpose of the present study was to examine the relationship of critical item content and assessment center ratings. Specifically, the study attempted to answer the question of whether targeting negative or positive items will effect overall assessment center ratings. It was hypothesized that participants being rated with negatively stated items would receive lower global, overall assessment center ratings than participants being rated with positively stated critical items. Additionally, since Cunningham (1982) has shown that volunteers are rated higher than nonvolunteers when global, subjective ratings are used, it was hypothesized that volunteers would be rated higher than nonvolunteers on these measures. Finally, it was hypothesized that behavioral measures would have less intercorrelation between dimensions than global ratings, demonstrating the decreased influence of halo effect in the more objective, behaviorally based rating system.

Method

Subjects

Three male volunteer, five male nonvolunteers, three female volunteer, and 11 female nonvolunteer subjects were recruited from the Business Administration Department at North Texas State University. Volunteers were recruited by

reading a description of a research project entitled "Management Assessment Center" in graduate classes. Individuals volunteering to participate placed their name, telephone number, and date most convenient for them on a sign-up sheet immediately after the announcement. Nonvolunteers participated as a requirement of a graduate level business course. Subjects were contacted in the order that they signed up until the sessions were full. Those who signed up for a particular date but are unable to participate were offered the next available date. All participants gave informed consent prior to participation. Subjects received verbal feedback outlining their performance as well as an occupational development plan.

Design

The design of the experiment was a 2 (positive vs. negative item content) x 2 (volunteer vs. nonvolunteer) multiple analysis of variance with global ratings of management potential and current management level as the dependent variables. One group of subjects was rated on a behavioral checklist with positively stated critical items while the other group was rated with a negatively stated behavioral checklist. Additionally, correlation matrices were computed showing the intercorrelations among assessment dimensions for global rating format and the behavioral checklist format. Since the intercorrelation of dimensions is an index of the halo effect,

this analysis provided data allowing a comparison of the levels of halo inherent in the two rating methods.

Materials

The Assessment Center Survey. The Assessment Center Survey consists of 20 Likert-type scales, fill-in-the-blank questions, and multiple-choice questions which assess the participant's knowledge of, attitude toward, and prior experience with assessment centers.

The Background Information Blank. The Background Information Blank is a questionnaire designed to gather information regarding educational background, major employment history, personal and family history, and life goals. Similar areas thought relevant to management success such as occupational successes and disappointments are also addressed.

In-Basket Exercise. The in-basket exercise is an exercise in which each participant is supplied with a packet of materials and asked to assume the hypothetical role of an employee who has suddenly been promoted to a management position. Individuals are asked to assume that it is Saturday morning, they must leave in one hour to catch a plane, the switchboard is disconnected, and they must take action on the items pending in their predecessor's in-basket. The in-basket materials contains 27 items, consisting of memos from the president, anonymous letters from hostile employees, bills which must be handled, requests for vacation leave, and other items. The gist of the in-basket exercise is to set up a

skeletal idea of a department which is in need of serious attention. Participants were asked to take action on each item and to record those actions in writing.

In-Basket Interview. The in-basket interview is an exercise in which an experimenter questions the participant about specific items in the in-basket. The experimenter agrees with the way the participant handles some items, questions some decisions, flatly disagrees with other decisions, and pushes for more discussion on a final item. The rater's questions and reactions are standardized on the assessor instructions for the in-basket interview.

The Compensation Committee. The Compensation Committee exercise is a leaderless, group exercise in which the participants are asked to continue in the role of the new manager. They are asked to assume that they have been appointed to a six person committee which dispenses discretionary salary increases to deserving employees. Data pertaining to six employees' initial and current salary, the date of their next regular merit review, their academic degrees, their last performance evaluation, and a letter of recommendation from a supervisor are given. The employees in the exercise are all mentioned in the in-basket exercise, making earlier information important for the participant's performance in the committee meeting. Each participant was pre-assigned an employee to represent. The task was to obtain at least $1000 for his or her employee, while at the same time helping the committee

make the best use of the $4000 allocated for such increases. They had one hour to divide the money, or no one receives a raise.

The Group Exercise Participant Report Form. The Group Exercise Participant Report Form requires the participants to rank order each other on dimensions such as contribution to the exercise and effectiveness in the group. Additionally, the participants are asked to indicate which group member is the quickest to understand the exercise, who is the leader, who is the strategiest, who disrupted the most, who is concerned with other's feelings, and who most attended to time. Finally, the participants are asked to rate their peers on how well they would perform in a group of the best people management could assemble for a group such as the one just completed. They are also asked to indicate how much money they were able to get for the employee that they represented in the exercise.

Policies and Procedures Committee. In the Policies and Procedures Committee exercise, the participants are asked to chair a three-person committee which is charged with the duty of developing policies and procedures addressing employees' appearance and behavior during breaks. Again, these problems stem from the in-basket materials, so information from that exercise is essential for adequate performance in this committee.

Job Interviews Exercise. The Job Interviews exercise is a two part exercise in which participants received a vitae and a letter of recommendation for two job applicants. During the first part of the exercise, they interview a person who is technically proficient, mature, and competent on the job. However, this applicant is described as somewhat shy and withdrawn and may be interpersonally abrasive with subordinates.

The second part of the exercise asks the participants to interview an applicant who is less experienced and technically proficient. However, this applicant has excellent interpersonal skills and is a quick learner. Participants may hire both, one, or neither of the applicants. Further, they may postpone their decision until later.

Secretarial Job Descriptions Committee. In the Secretarial Job Descriptions Committee exercise, the participants are asked to chair a three-person committee which meets to write concrete, unambiguous job descriptions, including duties and responsibilities, for secretarial staff. The secretarial pool is mentioned in the in-basket materials as a problem area in the department. The participants must discuss the problem and prepare a written outline of secretarial job descriptions.

The Oral Presentation. The oral presentation is a role-play exercise in which the participants are asked to review 21 statements about a computer system and then to try to sell

their supervisor on the idea of purchasing a particular system. During the 20 minute exercise, a confederate playing the supervisor listens to the presentation and then begins to make statements with sexual overtones. The confederate reactes to the participants' responses in a way that is obviously sexual harassment. In this exercise, as in every role play exercise, the confederates are graduate students trained by a professional actress. The confederates are randomly assigned to participants in each exercise, with the condition that each participant interacts with each confederate no more than once. Additionally, in the oral presentation, female participants interact with male confederates and male participants interact with female confederates.

Performance Appraisal Interview. During the performance appraisal interview, the participants are to conduct a standard employee-performance appraisal with a secretary who has been described in the in-basket materials as invaluable to the company but who is emotionally upset because she feels that she gets all the undesirable jobs, works harder than her peers, and makes less money. A confederate assumes the role of an irate secretary who screams, cries, moans, and does anything to get her way.

The Reorganization Plan Committee. The Reorganization Plan Committee is another three-person committee meeting. This committee is charged with reorganizing the division so as to make it more responsive to company needs. They are given a

budget of $10,000 with which to affect the change. The participants are asked to prepare a written report justifying any expenses and showing how the changes can make the system more responsive.

The Management Objective Plan. The management objective plan is an exercise which calls for both a written and an oral response. The participants are given a computer printout which plots the bugets of several divisions, including their own, against that division's productivity. All other divisions are producing more as their budget increase. However, in the participants' division, the budget is steadily climbing while the productivity is steadily decreasing. The second part of the printout plots each of several factors (number of complaints, turnover rate, percentage of employees working all day, and the employee evaluations) against productivity. All of the data in this section pertains to the participants' division. The data is analyzed, using scattergram plots, multiple regression, and several other statistical techniques. Thus, there are several ways of understanding the gist of the printout. Once they analyze the data, the participants are asked to prepare a "management by objective" (MBO) plan. They are required to make specific suggestions that will help increase production, insure a smooth running operation, and offer a system for the evaluation of their objectives. Finally, after the written MBO is prepared, they are asked to present the plan to representatives from the company's

Board of Directors. The Board was made up of confederates who asked questions based on the MBO, as well as prepared questions designed to make participants "think on their feet."

Background Interview. The background interview is an exercise in which participants are interviewed by the experimenters. It is the only exercise in which the participants drop the role of the hypothetical manager. The experimenters gather information about the participants' personal and employment histories that is relevant to their predicted success as managers.

The Self-Rating Form. The Self-Rating Form asks the participants to rate themselves on a seven-point, Likert-type scale on each of the 12 exercises. A rating of one is anchored with "poor," while a rating of seven corresponds to "very good."

The Gobal Rating Form. The Global Rating Form calls for the participants to rate themselves on a seven-point, Likert-type scales on each of 10 dimensions. The dimensions are leadership, written and oral communication, problem analysis, delegation, interpersonal skills, stress tolerance, organization and planning, functional knowledge, management level, and management potential. The scales are anchored at the extreme ends with short, descriptive phrases. Phrases on the high end of the scale are more positive, while the phrases on the low end are negative managerial characterisitcs. However, the ratings of management level and managerial potential

are reverse scored, with a low score indicating a high ability level and a high score indicating a low ability level.

The Exercise Evaluation Form. The Exercise Evaluation Form is a two-part credibility form. The first part consists of seven questions about the credibility of each exercise. The second part asks the participants to rate the credibility of the assessment center process as a whole. Ratings are given on a nine-point, Likert-type scale with anchors at the extreme ends.

The Sixteen Personality Factor Questionnaire (16PF). The 16PF is an objectively scored personality test. The test yields 16 primary factor scores: reserved vs outgoing (A), less intelligent vs. more intelligent (B), affected by feelings vs. emotionally stable (C), humble vs. assertive (E), sober vs. happy go lucky (F), expedient vs. conscientious (G), shy vs. venturesome (H), tough-minded vs. tender-minded (I), trusing vs. suspicious (L), practical vs. imaginative (M), forthright vs. shrewd (N), self-assured vs. apprehensive (O), conservative vs. experimenting (Q1), group dependent (Q3), and relaxed vs. tense (Q5). Alternate form reliability for this instrument is reported to be .70, and criterion validity coefficients areas high as .56 (Buros, 1975). A Sten score is obtained. Scores 1-3 and scores 8-10 are interpretable.

The Self-Evaluation Questionnaire. The Self-Evaluation Questionnaire is an objectively scored test of state and trait anxiety. Spielberger et al., (1979) report one month

test-retest reliability coefficients of .84 for males and .76 for females on the trait score and .33 for males and .16 for females on the state scores. Since state anxiety is expected to vary over time, the low reliability coefficients for this score are not to be considered problematic. Validity correlations with the Institute of Personality and Ability Testing Anxiety Scale (.77) and the Taylor Manifest Anxiety Scale (.83) are encouragingly high. The participant responds to short statements by checking a four-point Likert-type scale, ranging from "almost never" to "almost always." The answers are scored one through four, with some items being reverse scored because some statements are positive and some are negative. Raw scores are transformed into percentiles, as given in the test manual. Scores of 85th percentile and above on the State Anxiety section indicate that the respondent consciously perceives tension and apprehension. This is the type of anxiety expected in highly stressful situations such as the assessment center. A score of 85th percentile or higher on the Trait Anxiety portion means that the respondent has a tendency to be anxious over a wide variety of situations or events. These people usually exhibit anxiety as a general behavioral disposition.

The Rathus Assertiveness Schedule. The Rathus Assertiveness Schedule is an objectively scored instrument used to assess assertive behavior. Rathus (1973) reports a test-retest reliability coefficient of .78 and a split-half

reliability of .77. The author also reports validity correlations for the schedule with measures of how people respond in specific situations to be .70.

A score is given for each item, ranging from -3 to +3. A total score is obtained by adding the item scores, after changing the signs of reversed items. Scores of -90 to -25 for females and -90 to -15 for males are interpreted as indicating a passive, nonassertive, or shy person who holds back self-expressions and sometimes feels taken advantage of. Scores of -24 to +25 for females and -14 to +35 for males are interpreted as indicating a person who expresses himself or herself appropriately, is at ease handling business situations, speaks up when his or her rights are violated, and may enjoy a good argument. Scores of +26 to +90 for females and +36 to +90 for males are interpreted as indicating a person who is orally assertive to the point of being aggressive, has a tendency to brag, and frequently antagonizes others.

The Edwards Personal Preference Schedule. The achievement scale from the Edwards Personal Preference Schedule is an objectively scored instrument that indicates the respondent's need for achievement. Buros (1975) reports "impressive" Kuder-Richardson Formula 20 reliability coefficients, with a range from .65 to .95. However, Buros also states that the validity data for this test is "meager." Scoring is done by adding the number of responses checked. Raw scores are

converted into percentiles. Scores of 97th percentile and above are interpreted as "very high." Scores of 85th percentile to 96th percentile are interpreted as "high." Scores of 17th to 84th percentile are interpreted as "average." Scores of 4th to 16th percentile are interpreted as "low." Scores of 3rd percentile and below are interpreted as "very low."

The Incomplete Sentence Blank for Adults. The Incomplete Sentence Blank for Adults has been revised to have face validity for management students and working managers. The blank is scored projectively with dimensions such as areas of conflict, likes, dislikes, interests, and similar areas in mind. There is no specific scoring method for this technique.

Mach V Scale. The Mach V Scale was constructed to measure the extent to which one views and manipulates others for his or her own purposes. Items are classified into one of three groups: advocacy of manipulative tactics in interpersonal relations; unflattering view of the nature of humans as being weak, cowardly, and easily subject to pressure from others; and abstract morality.

The scale is composed of 20 items with three statements each. Respondents are to indicate which statement they most agree with and which statement they least agree with. Scores are assigned based on the combination of statements checked within each item. The possible range of scores is 20 to 140. The mean for males is 93.69, with a standard deviation of of 14.37; whereas the mean for females is 87.66, with a

standard deviation of 13.45. Christie and Geis (1970) report the item-whole correlation for all items to be .38. No validity data are reported.

Personal Attributes Inventory. The Personal Attributes Inventory is an objectively scored measure of masculinity, femininity, and androgeny. There are eight items corresponding to each of the three dimensions. High scores on the dimensions indicate that the respondent has those socially desirable characteristics that are stereotypically attributed to members of one or the other sex, or to both sexes. The items consist of a pair of characteristics with letters "A" through "E" between them. "A" corresponds to a rating of having very little of the trait, whereas "E" corresponds to having a lot of the trait. Spence (1978) reports Chronbach alphas for a sample of students to be .85 for masculinity, .82 for femininity, and .78 for androgeny. Also reported are validity correlations with the Bem Sex Role Inventory. These validity coefficients are .75 for males and .59 for females. The scale is scored by assigning the score of 1 to "A" response, 2 to a "B" response and so on, with "E" receiving a score of 4. The item scores are summed for each dimension for the total score on the dimension. Mean scores for college students are 21 (standard deviation of 4.5), for masculinity, 23 (standard deviation of 4.0) for femininity, and 15 (standard deviation of 4.0) for androgeny. High

scores are associated with socially desirably characteristics of the sex corresponding to the dimension.

Self-Monitoring Scale. The Self-Monitoring Scale is a 25-item, objectively scored instrument that measures the extent to which individuals monitor (observe and control) their expressive behavior and self-presentations. It also measures the person's sensitivity to the expression and self-presentation of others and the ability to use these cues as guidelines for managing their own expressive behavior. The scale is scored by adding one point for each "correct" response. The total score is the total number of points. A score higher than 15 is considered "high," and a score of nine or less is considered "low."

Synder (1974) reports the Kuder-Richardson 20 reliability to be .70 and the test-retest reliability to be .83. The author also cites a validity study wherein the subjects' score on the Self-Monitoring Scale correlated closely ($r = .45$) with peer ratings of that person's self-monitoring ability. This article also reported that actors' mean score was significantly higher than that of college-student controls ($t = 8.27$, $df = 555$, $p < .001$). Persons scoring high on the scale are good at learning what is socially appropriate in new situations, have good self-control over emotional expression, and can effectively use this ability to create desirable impressions.

Levenson's Locus of Control. Levenson's Locus of Control is an objectively scored instrument that measures the

respondent's belief that his destiny is controlled by chance, powerful others, or himself or herself. The scale consists of 24 items in Likert-type format, representing the three dimensions of internal control, control by powerful others, and control by chance. The range on each scale is from 0 to 43. The mean score for Internal (I) control is 35.48, for Powerful Others (P) is 16.65, and for Chance (C) is 13.94.

High scores on the internal dimension suggest that the respondent feels in control of his or her life and is usually involved in several activities designed to control some aspect of his or her life. High scores on the Powerful Others scale reflect a belief that other people are in control of the respondent's life. High scores on the Chance scale reflect the belief that fate controls the life of the respondent. The individuals are unlikely to engage in activities to control their own lives. Levenson (1972) reports no reliability coefficients but states that "retesting indicated that the I,P, and C scales had a high internal consistency and were not correlated with a measure of social desirability" (p. 262). Levenson's validity study compares high versus low internals on involvement in activities designed to control their lives. The group differences were significant ($F = 5.76$, $p < .02$), establishing validity for this dimension.

Adjective Checklist. The Adjective Checklist consists of 300 adjectives commonly used to describe attributes of a

person. Subjects are instructed to check those which they consider self-descriptive. The scale may be completed in 10 or 15 minutes, even by unsophisticated subjects, and arouses little resistance and anxiety. Currently, there are 24 experimental scales available for use with the Adjective Checklist. They are total number of adjectives checked, defensiveness, number of unfavorable adjectives checked, self-confidence, self-control, lability, personal adjustment, achievement, dominance, endurance, order, intraception, nurturance, affiliation, heterosexuality, exhibition, autonomy, aggression, change, succorance, abasement, deference, and counseling readiness.

The Adjective Checklist is computer scored. The score for each scale is determined by subtracting the number of contra-indicative adjectives (for that scale) from the number of indicative adjectives checked and then converting to a standard score according to sex and total number of adjectives checked.

Test-retest reliabilities for the scale cluster at .64. This low reliability indicates that the Adjective Checklist is subject to short-term fluctuations in mood. The authors found that reliability coefficients are significantly related to certain checked adjectives, so test-retest stability is a characteristic of the respondent as well as the test (Gough & Hielbum, 1965).

Gough and Heilbum (1965) report "considerable construct validity" in a wide variety of studies. These studies included prediction of college dropout, termination of counseling, and architectural creativity.

Repression-Sensitization Scale. Defense style was measured using Byrne's (1961) Repression-Sensitization (R-S) Scale, which consists of 156 Minnesota Multiphasic Personality Inventory (MMPI) items. The participants are instructed to mark true or false to each item as it pertains to their personality. The scale is scored by crediting one point for each answer as given in the Byrne article. Split-half and test-retest reliabilities of .88 were reported. Correlations with Ullman's Facilitation-Inhibition Scale were reported as -.75 for a group of students and -.94 for male psychiatric patients. Validity coefficients of .62 and .55 were also reported from correlations with Worchel's Self-Activity Inventory for two student samples.

The Interview Guide/Checklist. The Interview Guide/Checklist is a structured interview outline which covers such areas as early years and family history, education and training, work experiences, health and disability, economic conditions, personality characteristics, problem behavior, current interests, current activities, and personal goals. The form is useful in semistructured interview situations.

The In-Basket Score Sheet. The In-Basket Score Sheet is a form for objectifying the scoring of the in-basket

exercise. All of the items in the in-basket materials are listed and blanks are provided in which to code the participant's actions on each item.

The Assessor Instructions for the In-Basket Interview. The assessor instructions for the in-basekt interview is a structured interview format for use during the in-basket interview. Topics covered include general impressions and general problem areas of the division, as well as the relationships between the problem areas. Additionally, specific items from the in-basket materials are addressed. The interviewer is directed to agree with the way some items were handled, challenge the response to other items, and push for more discussion on a final item.

The MBO Questions Form. The MBO Questions Form lists eight questions designed to assess the participant's understanding of the day's activities. The questions call for information gained throughout the entire day's activities. Additionally, there are questions that require the participant to relate the hypothetical company's concerns to current issues in the business world. Finally, there are irrelevant and meaningless questions designed to assess the participant's aibility to screen out distractions in an appropriate manner.

The Positive Behavioral Checklist. The Positive Behavioral Checklist is a behaviorally based rating form. Eight dimensions are measured (leadership, problem analysis, communication, delegation, interpersonal skills, organization and

planning, stress tolerance, and functional knowledge). Under each dimension heading are seven operationally defined behaviors that reflect some attribute of the dimension. Each critical item is stated in the positive direction. Scoring is done by checking each attribute observed in each assessment center exercise. Then, a cumulative frequency for each dimension across exercises is computed, giving the summary scores for each dimension. Finally, the summary score of the dimensions are added, giving a total behavioral score.

The Negative Behavioral Checklist. The Negative Behavioral Checklist is a behaviorally based rating form. Eight dimensions are measured (leadership, problem analysis, communication, delegation, interpersonal skills, organization and planning, stress tolerance, and functional knowlgdge). Under each dimension heading are seven operationally defined behaviors that reflect some attribute of the dimension. Each critical item is stated in the negative direction. Scoring is done by checking each attribute observed in each assessment center exercise. Then, a cumulative frequency for each dimension across exercises is computed, giving the summary score for each dimension. Finally, the summary score for the dimensions are added, giving a total behavioral score.

The Descriptions Form. The Descriptions Form is similar to the Behavioral Checklist, in that dimension are listed (leadership, problem analysis, communication, delegation, organization and planning, stress tolerance, and interpersonal

skills), and several items are included under each category. However, on this form, the items are adjectives which describe various aspects of the corresponding dimension. No formal scoring is done.

The Global Rating Form. The Global Rating Form calls for the experimenters to rate the participants on a seven-point, Likert-type scales on each of 10 dimensions. The dimensions are leadership, written and oral communication, problem analysis, delegation, interpersonal skills, stress tolerance, organization and planning, functional knowledge, management level, and management potential. The scales are anchored at the extreme ends with short, descriptive phrases. Phrases on the high end of the scale are more positive, whereas the phrases on the low end are negative managerial characteristics.

The Self-Rating Form. The Self-Rating Form calls for the experimenters to rate the participants on seven-point, Likert-type scales on each of the 11 exercises. A rating of one is anchored with "poor," whereas a rating of seven corresponds to "very good."

Generic Job Descriptions Table. This table lists generic job descriptions for positions which are common to most companies. The positions are President, Executive Vice-President, Staff Vice-President, Operating Vice-President, Department Manager, and Group Supervisors. Descriptions are behaviorally oriented. The descriptions are used to assist in identifying

management levels that the participants are suited for immediately, as well as the level that their particular strengths and weaknesses make them ideally suited for.

The Development Planning Form. The Development Planning Form lists 25 possible training and development methods. Each method is targeted to address specific deficiencies in a participant's performance. The specific skills and abilities that are expected to accrue from participation in the training method are given.

Experimenters

The experimenters were 16 graduate students in business and psychology. The first step of the training process for the experimenters was to have them decide upon the dimensions to be assessed and then to operationalize those dimensions into observable behaviors. The dimensions were compiled from those most often reported in the literature. All behaviors were stated positively. Next, the experimenters practiced rating during pilot runs. Finally, the trainers co-rated with each experimenter during an acutal assessment center run, so as to be able to calculate interrater reliability. Each experimenter was trained to at least .85 reliability. Additionally, intermittent, random reliability checks were done during assessment center runs to control for observer drift.

After the raters were trained using the positive checklist, a negative checklist was devised. Raters again trained and practice rated using these negatively stated checklist items.

Raters, trained on both the positive and negative behavioral checklist, were randomly assigned to rate with one or the other form during assessment center runs. This assignment was randomized again each time the center ran.

Procedure

The participants signed up as a requirement of their graduate business class or in response to a call for volunteers. They were later called, in the order that they signed up, and were asked to participate in an upcoming assessment center run. Assessment center runs were scheduled on Saturdays during the school year.

Participants were randomly assigned to be rated with the Positive or Negative Behavioral Checklist. One half of the participants in each assessment center run were rated with each checklist.

At 7:30 a.m., on the day of the assessment center, the participants gathered in a large room for Orientation. First the participants completed an informed consent form, a background information form and some pre-center psychological inventories. Then, the layout of the building was described and each participant was assigned an office. They were told to assume the role of a newly assigned manager whose department is sorely in need of attention. Participant's schedules represented their "desk calander," showing all the appointments and tasks that were required for the day. They were told that their packet of psychological inventories must be

completed by the end of the day. Since it takes approximately two hours to complete the forms and the participants have only one hour of "free time," they must work the rest of the paper work into their busy schedule. The specific psychological inventories were chosen by a clinical psychologist with training in management assessment. The battery represents many tests that are commonly used in such an assessment as well as some inventories that were included for experimental evaluation. At the end of Orientation, the importance of maintaining the role of the hypothetical manager was again stressed as well as the importance of time management skills in making appointments, being prepared for exercises, and completing the psychological inventories.

After a few minutes for coffee, the in-basket exercise began, lasting from 8:15 to 9:15. At the end of the hour, the participants turned in these materials to a central office.

After turning in the in-basekt materials, the participants gathered in a conference room for the Compensation Committee. This exercise also lasted one hour. This was the first exercise in which the participants were rated, using the Behavioral Checklist.

At 10:15, after the Compensation Committee, the participants split into three person committees for the Policies and Procedures Committee exercise. During this 30 minute meeting, one of the three participants was appointed chairperson. Only the chairperson was rated.

The next exercise, lasting from 10:45 to 11:15, was the background interview. The interviewers had previously reviewed the participant's Background Information Blank. Interviews are semistructured and based on the Interview Guide/Checklist.

From 11:15 to 11:45 was the first part of the Job Interview Exercise. A male or female confederate played the role of the job applicant.

After the interview, lunch was served from 11:45 to 12:30. The participants were free to use this "unstructured" time anyway that they wished. During this time they ate, visited with each other, and/or worked on the psychological inventories. Additionally, they could have used this time to prepare for the next exercise.

After lunch, the participants gathered in a conference room for the Secretarial Job Descriptions Committee. During this exercise, one of the participants who did not chair the Policies and Procedures Committee was appointed to act as chairperson. Again, only the chairperson was rated. The exercise lasted from 12:30 until 1:00.

While the participants had free time, the raters were scoring the in-basket items in preparation for the in-basket interview. The interview was conducted from 1:00 until 1:30, following the Assessor Instructions for the In-Basket Interview. The participants were again rated on the Behavioral Checklist.

After this committee meeting, the participants had a free period, from 1:30 to 2:00. At 2:00, the Oral Presentation began. This is a 20 minute exercise. During this exercise, the experimenters rated both the participant's ability to present a coherent sales pitch for the computer system and the reaction to the confederate supervisor's sexual comments.

After another free time from 2:20 to 3:00, the Performance Appraisal Interview began. During this exercise, the experimenters rated the participant's reaction to the confederate's overt display of violent emotion, as well as their ability to stay task oriented and complete the performance appraisal.

At 3:30, after another 10 minute free period, the other part of the Job Interview exercise began. During this 30 minute exercise, the participant interviewed the other applicant for the secretarial pool. As in the previous job interview exercise, confederates could have been male or female.

Next, the last of the three person committees, the Reorganization Plan Committee, met. Participants who had not chaired a three person committee were appointed to act as chairperson. As before, only the chairpersons were rated. The committee meeting lasted from 4:00 until 4:30.

After another short free time period, the participants went to a conference room for the Management by Objective Plan presentation to the "Board of Directors." The Board is made

up of confederates who ask questions based on the presentation as well as the MBO Questions Form.

After the MBO, the participants were given a packet of post-center inventories. These inventories are designed to detect any personality changes over the day, rate the participants' concept of their own performance, and they evaluate the credibility of each exercise. Participants were then debriefed and were free to leave. Most participants finished the assessment center by 5:30.

The schedules were arranged so that each experimenter rated each of the six participants in at least one exercise. During the week after the assessment center run, the experimenters analyzed their data and wrote reports on the participants. The chief investigator then reviewed the raw data and supervised the experimenters on the report. Appointments were then made with the participants for individual, verbal feedback. This feedback included specific information on all of the evaluated dimensions, as well as information on suggested remedial actions for identified weaknesses.

Results

The dependent measures for the study were the mean of the experimenter's ratings of management potential and level, as well as the absolute value of the difference scores resulting from subtracting rater's estimates of management potential from participant's self ratings of management potential. The score on management potential is a rating of the highest management

level that the participant is likely to achieve. The ratings are done on a seven-point, Likert-type scale. The score on management level is a rating of the current management level that is best suited to the participant's ability. The ratings are done on a seven-point, Likert-type scale. The difference scores reflect the discrepancy between the participant's and the raters' estimates of the participant's management potential. Both the participant's and the rater's ratings are done on seven-point, Likert-type scales. On all Likert-type scales a lower score indicated better performance.

Critical Item Content by Volunteer Status MANOVA

A 2 (positive vs. negative item content) x 2 (volunteer vs. nonvolunteer) multivariate anlaysis of variance was prepared in which the dependent measures were combined to yield a simple F value. The mean and standard deviation for each dependent measure are listed in Table 1 (Appendix D). The analysis reveals a significant main effect for volunteer status, F (3,12) = 4.39, $p < .026$. Because this finding was significant, univariate analysis of variance were performed on three dependent measures. These analyses are summarized below.

Level. A 2 x 2 (item content x volunteer status) analysis on the ratings of management level reveals a significant main effect for volunteer status, F (1,14) = 12.63, $p < .003$. The volunteers were rated as better on this measure. No significant effect is found for item content or for the item

content x volunteer status interaction. Thus, the hypothesis that volunteers would score better than nonvolunteers were supported. However, the hypothesis that participants rated with positive critical items would score better than those rated with negative items was not supported.

Potential. A 2 x 2 (item content x volunteer status) analysis on the rating of management poetntial reveals a significant main effect for volunteer status, F (1,14) = 5.95, $p < .029$. Volunteers were rated as better on this measure. No significant effect is found for item content or for the item content x volunteer status interaction. Thus, the hypothesis that volunteers would score better than nonvolunteers was supported. However, the hypothesis that participants rated with positive critical items would score better than those rated with negative items was not supported.

Difference Scores. A 2 x 2 (item content x volunteer status) analysis on the difference scores reveals no significant main effects. Additionally, the item content x volunteer status interaction does not reach significance. Thus, neither of the hypothesis relating to volunteer status or critical items was supported.

Critical Item Content by Sex MANOVA

In order to determine if there were difference based on the participants' sex, a 2 (positive vs. negative critical item) x 2 (male vs. female) multivariate analysis of variance was calculated. The mean and standard deviation for each

dependent measure are listed in Table 2 (Appendix E). The analysis reveals no significant main effect. Additionally, the sex by item content interaction was not significant.

Critical Item Content by Rater Experience MANOVA

It was thought that previous experience rating assessment center participants may have had some influence on the rater's susceptability to cognitive set and the resultant rating bais. As a supplimentary analysis to isolate the effects of the raters' previous experience with assessment center ratings, a 2 (high vs. low experience) x 2 (positive vs. negative critical item checklist) multivariate analysis of variance was calculated. The dependent measures were the mean global ratings of potential and level given by each rater. The mean and standard deviation for each dependent measure is listed in Table 3 (Appendix F). The analysis reveals no significant main effect or interacting effect.

Correlation Matrices

Following these analyses, global ratings of each of the eight management dimensions and summary scores from the Behavioral Checklist's eight dimensions were used to calculate two correlation matrices. These ratings are estimates of the participant's competence in the eight assessed dimensions. The global ratings are done on seven-point, Likert-type scales. The behavioral rating scores are the sum of all items checked within a particular dimension, across all simulation exercises. This analysis allows a comparison

of the intercorrelations of the assessed dimensions, and hence, the scale's susceptability to halo effect. These matrices are presented in Tables 4 and 5 (Appendices G and H). The average correlation for each matrix was then computed. The average correlation for the behavioral rating is 0.58. The average correlation for the global rating is .068. Thus, the hypothesis that behavioral ratings are less susceptable to halo effect than are global ratings was supported.

Discussion

Hypothesis 1: Critical Item Effect

Srull and Wyer (1980) found that making different critical items, or trait items, available to a rater can influence subsequent ratings of target persons. This fact led to the hypothesis that those participants rated with positive critical items would receive higher ratings than participants who were rated with negative critical items. Studies reported by other authors extended this finding to several populations (Huguenard et at., 1970; Nisbett & Wilson, 1977) and offered rationals for the effect (Darley & Fazio, 1980; Snyder & Swann, 1978). However, the results of the present experiment do not support this hypothesis.

Perhaps Srull and Wyer's (1980) "priming" effect can account for the lack of significant findings for the item content variable in this study. All raters were originally trained using positively stated behavioral checklists. Only

after becoming thoroughly familiar with this checklist did they develop and train on the negatively stated checklist. Therefore, it is possible that using the positively stated items first overshadowed the influence of the actual critical items used to evaluate the participants. In other words, perhaps the positively stated critical items were "primed" into the rater's cognitive set and that the influence of this "priming" was powerful enough to negate the influence of the checklist used during the assessment.

Research on primacy and recency effects has direct bearing on this issue. Leach (1974) has shown that the instructions given prior to rating a target person determines whether earlier or later information about the target person is most influential when forming a subsequent rating. When told to rate subjects after presenting two sets of information, the final ratings show strong recency effects. Thus, the most recent material presented seems to hold greater sway in influencing raters' judgements. However, if the raters' are explicitly told that both pieces of information are about the same person and raters are urged to form a unified picture of the target person, more ambiguous, less extreme ratings emerge. It seems that under these instructions the recency effect disappears and the ratings are more consistent with an integrated picture of the person.

Mayo and Crockett (1964) have shown that the rater's cognitive complexity interacts with the order of information

presentation when making global judgements about target persons. Cognitive complexity is conceived of as the degree to which a person can make differentiations between persons. Someone who shows a facile ability to differentiate among those whom he comes in contact with is thought of as cognitively complex. Someone who has a poor sense of differentiation is said to be cognitively simple. Cognitively complex raters are thought to make more inferences about target persons than do cognitively simple raters. Additionally, cognitively complex raters are more likely to ascribe positive and negative traits to the target and to give multiple meanings to target behavior. The authors hypothesized that if given two sets of disparate information, raters low in cognitive complexity will reject the first and accept the second set of information. In their study, raters high in cognitive complexity, on the other hand, were expected to use their broader range of inferences to help them assimilate the material into an integrated whole. Raters exhibiting low cognitive complexity did, in fact, show extreme receny effects. They were strongly influenced by material presented later in sequence. However, those raters who demonstrated high cognitive complexity tended to integrate information to form less extreme, more ambivalent judgements about the target persons.

The variables discussed by Leach (1974) and Mayo and Crockett (1964) were probably influential in the present study. First, raters were accutely aware that all information

collected on each participant would be integrated into a global rating as well as an assessment center report. This should promote the formation of a unified picture of the target person. Additionally, the raters were highly trained and competent at assessing managerial behaviors and many of them were also trained as psychologists. They were trained to recognize a variety of management behaviors and to analyze the functional utility of clusters of behaviors rather than single behaviors. Any single behavior, then, could be construed as positive or negative, depending on the other attendant behaviors. Given such an elaborate system for differentiating between managers, it seems safe to conclude that they had developed a rather sophisticated cognitive structure. Again, this factor facilitates the formation of a unified impression of the target person which is based on information gathered at different times. These two factors, then, probably account for the negligable recency effects. When using the negatively stated checklist, the resultant ratings integrated information which was based on both sets of critical items. Since the negatively stated items were so similar to the positively stated ones, the raters had little difficulty remembering the more positive trait categories.

Hypothesis 2: Volunteer Status Effect

The results of the present experiment support the hypothesis that volunteers will be rated as having performed

better than nonvoluntters on global ratings of management performance. This result supports Cunningham's (1982) findings. In the present study, differences relating to volunteer status were found on both the ratings of management potential and management level. However, the discrepancy between raters' and participant's estimates of management potential was not significant, indicating that both groups formed equally accurate conceptions of what ratings they would receive. Since volunteers were rated as having performed better than nonvolunteers, the finding suggests that nonvolunteers are more likely than volunteers to perceive their performance as being rated as poor.

These findings have important practical and ethical implications. First, the findings lend credibility to Cunningham's (1982) contention that global ratings for management assessment centers should be abandoned in favor of more objective, behavioral measures. He found that the behavioral measures do not discriminate between volunteers and nonvolunteers. Thus, the behavioral ratings are seen as a more fair assessment method in those centers accepting participants who volunteer as well as those who are forced to attend. Additionally, these results suggest that nonvolunteers may leave the assessment center feeling less positive about their performance than do volunteers. Such negative feelings may subsequently effect their self-esteem and self-confidence. In short, being forced to participate in an

assessment center may, in fact, be deleterious to some participants.

Since volunteer status effect is an important consideration when devising participant selection methods, further research is needed. Only by illuminating the important determinates of performance can the personnel officer be confident that the assessment is fair and accurate.

Hypothesis 3: Halo Effect

The finding that the intercorrelations between management dimensions is lower for behavioral ratings than for global ratings suggests that behavioral ratings are less influenced by halo effects. Since halo is considered "error variance," this advantageous reduction of halo again supports the contention that global ratings of management performance should be abandoned in favor of behavioral ratings (Cunningham, 1982). Since globally rated dimensions are highly correlated, it appears that raters did rely on central traits when making assessment center ratings (Cooper, 198; Jacobs & Zedeck, 1980).

These results suggest that specifically denoting the critical behaviors can counteract the tendency to rely on central traits when making judgements. Rater's attention is diverted from central traits to specific behaviors, so it becomes increasingly less likely that ambiguous behaviors will be interpreted as being consistent with prematurely made judgements. Behaviors are simply less likely to be perceived as ambiguous, the behavior either occurs or it is

absent. In contrast to the typical process of impression formation (Snyder & Swann, 1978), confirmatory behaviors will not be remembered as occuring more frequently. A frequency count of targeted behaviors will be directly available to the rater when making subsequent judgements. Additionally, "impression-maintenance attributional bias" (Darley & Fazio, 1980) will be decreased if behavioral measures are used. Since all the rated behaviors will be reviewed later with the aim of integrating the information, a more accurate, multi-dimensional picture of each participant should emerge (Mayo & Crockett, 1964).

Future Research

Since primacy and recency effects confound research wherein raters are trained on both sets of critical items, future researchers should train raters on positively or negatively stated items only and then compare ratings for randomly assigned participants. Another, more subtle, test would be to counterbalance training items such that some raters are trained on positively stated items first while other raters are trained on negatively stated items first. These designs, used together, would allow evaluation on both the training and rating components of the "priming" effects of cognitive set. If these studies show a biasing effect, perhaps combining positive and negative critical items will result in a more accurate picture of the participant.

Validity studies comparing positive, negative, and mixed critical item behavioral ratings are indicated.

Ethical considerations demand the assessment of the volunteer status variable of management assessment performance. Future research designed to assess the self-esteem and self-confidence of volunteers and nonvolunteers before and after the assessment center as well as after receiving feedback is indicated. If, as the results of this study indicate, nonvolunteers have lowered self-image after participation in an assessment center, some intervention is needed. Perhaps a debriefing session immediately after completion of the center or a thorough feedback session shortly after participation is in order.

Summary and Conclusions

Traditionally used global ratings are influenced by the cognitive set of the rater. Although this study found no significant differences in the ratings of participants rated using positive as opposed to negative critical items, the primacy effect of having trained the raters on the positive items first probably accounts for the finding. Prior research has shown that instructions which explicitly require cognitively complex raters to integrate disparate material will result in ratings based on both sets of data. The subsequent ratings will more accurately reflect the characteristics of the target person. Since similar instructions were in effect and since trained raters have complex cognitive

structures for evaluation of managers, these factors probably accounted for the observation of a negligable recency effect.

Global ratings were also shown to be sensitive to the volunteer status of the participants. Variables accounting for volunteer differences are directly assessed by the behavioral assessment (Cunningham, 1982) or have been shown to be irrelevant (Bray, 1982) to managerial success. Therefore, the use of global ratings should be discouraged because they discriminate along dimensions considered irrelevant to managerial success.

Global ratings are also more subject to halo effects than are behavioral ratings. Participant behavior is more likely to be conceived of as relating to some central traits when using global as opposed to behavioral ratings. Trait categories are then used to classify all subsequent behaviors. Such a process will cause participant behavior to change due to confirmatory seeking behaviors of the rater. High intercorrelations among globally rated dimension suggest that the ratings are heavily influenced by such artifacts.

A more advanced technology for assessing managerial performance and potential was offered. This technology relies on behavioral ratings which take place during an integrated series of management simulation exercises. The rating format reduced halo effect and sensitivity to irrelevant dimensions of performance such as "likeability" while the integration of the exercises allows assessment of often ignored but

important dimensions such as time management and the ability to use previously learned material when solving a subsequent problem. These procedures have also been shown to increase content validity of the assessment, to increase objectivity, and to reduce the importance of being interpersonally outgoing. Hence, these new procedures have clear advantages over traditional assessment center methods.

This study suggests areas of interest for future researchers. Designs which control for primacy and recency effects of training on positively or negatively stated critical items will help determine the effects of "priming" on subsequent rating. Validity studies comparing positive, negative, and mixed critical items will illuminate the best assessment methodology for future use. The effects of volunteer status on the participants' self-esteem and self-confidence as well as research into possible remedies is also indicated.

Appendix A

I _____, freely consent to be a participant in the project entitled "Assessment Center Approach to Managerial Selection" to be conducted at N.T.S.U. during the period of September 1, 1981 through May 31, 1982 with Alvin Smith and Donna Ledgerwood as the principal investigators. The procedures to be followed and their purpose, including identifying any procedures which are experimental, have been explained to me, and I understand them. They are: I understand that I will be asked to complete several questionnaires. I will be involved in several role play exercises and group exercises. I understand that the purpose of the project is to examine managerial skills through the use of situational exercises and questionnaires. I understand that after participation, I will be fully informed of the results.

The attendant discomforts and risks reasonably to be expected by my participation in this project have been explained to me, and I understand that they might be as follows: I may become tired or fatigued due to the length of the procedures. I might find some questionnaire items upsetting or confusing. I understand that I should feel free to discuss any feelings I have due to my participation and that I may stop at any time.

Any benefits reasonably to be expected from my participation have been explained to me and are as follows: I understand that I will be given individual feedback regard critical areas related to successful management; I will become familiar with the assessment center approach to selection; I will also receive recommendations regarding managerial development.

I understand that this consent and data may be withdrawn at any time without prejudice. I have been given the right to ask and have answered any inquiry concerning the foregoing. Questions, if any, have been answered to my satisfaction, I have read and understood the foregoing.

_____ _____

Signature Date

(check one)
Are you a volunteer _____ or is participation a course requirement _____?

If participation is a course requirement, would you have volunteered anyway? _____

Appendix B

Positive Behavioral Checklist

RATER EXERCISE PARTICIPANTS

I. Leadership
   1. Makes desision and announces it
   2. "Sells" decision
   3. Presents ideas and invites questions
   4. Presents tentative decision subject to change
   5. Presents problem, get suggestions, make decision
   6. Defines limits, asks group to make decision
   7. Permits subordinates to function within limits defined by superior

II. Problem analysis
   1. Identifies global problems
   2. Identifies specific problems
   3. Identifies organization/structional/procedural problems
   4. Identifies human factor problems
   5. Identifies relationship between problems
   6. Identifies alternatives, solutions, contingencies
   7. Demonstrates flexibility in viewing problems

III. Communication
   1. Voice clear with good delivery and articulation
   2. To the point without omitting information
   3. Easy to follow directions or grasp content
   4. Presentation systematic; organized; structured
   5. Evidence of human component considered (sensitive)
   6. Writing easy to read
   7. Communication stimulates discussion

IV. Delegation
   1. Assigns task (specific work)
   2. Assigns authority, responsibility, decision-making
   3. Assigns to three or more individuals
   4. Assignments are specifically clear
   5. Assignments are accompanied by dealiness or time frame
   6. Assignments accompanied by sensitivity (please, thanks, etc.)
   7. Assignments are appropriate (right work or right people)

V. Interpersonal Skills
   1. Attends to others while they speak (no interruptions)
   2. Reinforces others expressions
   3. Responds to others feelings

Appendix B--Continued

4. Looked to by others for affiliation
5. Demonstrates range of behavior and affect
6. Responds to others sincerely
7. Assists others in structuring, organizing, clarifying

VI. Organization and Planning
1. Arranges materials orderly before processing
2. Separates relevant and irrelvant information
3. schedules appointments at appropriate times
4. Clearly states objectives in plans
5. Outlines procedures to carry-out plans
6. Uses calendar (5 items scheduled)
7. Presentation is orderly and logical

VII. Stress Tolerance
1. Tolerance ambiguity without decrease in performance level
2. Manages large amounts of information without decrease in performance
3. No signs of excessive nervousness when challenged
4. Makes all appointments within five minutes of appointment time
5. Maintains position when challenged
6. Does not become disorganized under pressure
7. Finishes exercise

VIII. Functional knowledge
1. Demonstrates alternate ways of motivating employees
2. Reads and understands instructional materials
3. Working knowledge of functions of various divisions
4. Functional use of basic business terms
5. Stays within limits of position
6. Actively contributes to the content of the exercise
7. Considers the political ramifications of their actions

Appendix C

Negative Behavioral Checklist

RATER EXERCISE PARTICIPANTS

I. Leadership
1. Relies on others to make decisions
2. No attempt to "sell" decision
3. Presents ideas without inviting questions
4. Considers decisions final, not subject to change
5. Makes decisions without presenting problem and getting suggestions
6. Does not define limits, ask group to make decisions
7. Arbitrarily limits functioning of subordinates

II. Problem Analysis
1. Ignores global problems
2. Ignores specific problems
3. Ignores organizational/structural/procedural problems
4. Ignores human factor problems
5. Fails to see relationship among problems
6. Identifies only one solution to problems
7. Is rigid when viewing problems

III. Communication
1. Verbal delivery unclear, inarticulate
2. Not to the point, omits information
3. Hard to follow directions or grasp content
4. Presentation disorganized, poorly structured
5. Insensitive to human component
6. Writing hard to read
7. Communications do not stimulate discussion

IV. Delegation
1. Unwilling to assign specific tasks
2. Unwilling to assign authority, responsibility, decision-making
3. Assigns to less than three individuals
4. Assignments are unclear
5. No deadline or timeframe accompanying assignments
6. No evidence of sensitivity when delegating
7. Assignments are inappropriate

V. Interpersonal Skills
1. Interrupts other
2. Ignores or depreciates others for affiliation
3. Ignores other's feelings
4. Not looked to by others for affiliation
5. Limited range of behavior and affect
6. Responses to others are insincere

Appendix C--Continued

7. Does not assist others in structuring/organizing/clarifying

VI. Organization and Planning
1. Processes information before orderly arranging items
2. Does not distinguish relevant from irrelevant information
3. Appointments scheduled at inappropriate times
4. Objectives not clearly stated
5. Procedures to carry out plan out outlined clearly
6. Does not use calendar (less than 5 items scheduled)
7. Unorganized or illogical presentation

VII. Stress Tolerance
1. Ambiguity causes decreased performance
2. Large amounts of information interfere with performance
3. Shows signs of nervousness when challenged
4. Over 5 minutes late for appointments
5. Reverses position when challenged
6. Becomes disorganized under pressure
7. Gets off-task or leaves before finishing exercise

VIII. Functional Knowledge
1. Relies on one method of motivating employees
2. Unfamiliar with content of instruction materials
3. No knowledge of function of various divisions demonstrated
4. Misunderstands basic business terms
5. Exceeds limits of position
6. Contributions irrelevant to content of the exercise
7. Ignores political ramification of their actions

Appendix D

Table 1

Means and Standard Deviations for Potential, Level, and Difference Scores Listed by Checklist and Volunteer Status

Checklist	Volunteer Status	Potential		Level		Difference	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Positive	Volunteer	2.67	1.48	4.38	1.38	0.92	1.06
Negative	Volunteer	2.33	1.41	4.50	0.71	0.33	0.00
Positive	Nonvolunteer	4.04	1.12	6.04	0.45	2.02	1.39
Negative	Nonvolunteer	4.06	0.94	5.82	0.71	1.80	0.69
Entire Sample		3.64	1.26	5.52	1.01	1.61	1.15

Appendix E

Table 2

Mean and Standard Deviations for Potential, Level, and Difference Scores Listed by Checklist and Sex

Checklist	Sex	Potential		Level		Difference	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Positive	M	2.77	1.64	4.90	1.52	1.37	1.66
Negative	M	2.94	1.42	5.00	1.00	1.28	1.11
Positive	F	4.16	0.81	5.90	0.63	1.83	1.13
Negative	F	4.04	1.03	5.80	0.76	1.83	0.65
Entire Sample		3.64	1.26	5.51	1.01	1.61	1.15

Appendix F

Table 3

Means and Standard Deviations for Level and Potential Listed by Experience of Rater and Checklist

Experience Level	Checklist	Level		Potential	
		$\bar{X}$	SD	$\bar{X}$	SD
High	Positive	5.27	0.95	3.23	0.50
High	Negative	5.53	0.31	4.07	1.28
Low	Positive	5.73	0.64	3.83	0.15
Low	Negative	5.60	0.20	4.03	0.57
Entire Sample		5.53	0.54	3.79	0.73

Appendix G

Table 4

Correlation Coefficients Between Behaviorally Rated Dimensions

	Leadership	Problem Analysis	Communication	Delegation	Inter-personal Skills	Organization and Planning	Stress Tolerance	Functional Knowledge
Leadership	1.0000	0.4930	0.4680	0.4711	0.3908	0.6272	0.4710	0.5656
Problem Analysis	0.4930	1.0000	0.6388	0.3598	0.5602	0.6548	0.6119	0.5335
Communi-cation	0.4680	0.6388	1.0000	0.3801	0.9009	0.7879	0.9089	0.7892
Delegation	0.4711	0.3498	0.3801	1.0000	0.4456	0.2334	0.4535	0.2479
Interpersonal Skills	0.3908	0.4502	0.9009	0.4456	1.0000	0.6309	0.8763	0.6788
Organization and Planning	0.6272	0.6548	0.7879	0.2334	0.6309	1.000	0.7148	0.6913
Stress Tolerance	0.4710	0.6119	0.9089	0.4535	0.8763	0.7148	1.000	0.6687
Functional Knowledge	0.5656	0.5335	0.7892	0.2479	0.6788	0.6913	0.6687	1.0000

Appendix H

Table 5

Correlational Coefficients between Globally Rated Dimensions

	Leadership	Communication	Problem Analysis	Delegation	Interpersonal Skills	Stress Tolerance	Organization and Planning	Functional Knowledge
Leadership	1.0000	0.8555	0.8837	0.8045	0.5832	0.6252	0.8461	0.5943
Communication	0.8555	1.0000	0.7842	0.6718	0.6440	0.4972	0.8709	0.5206
Problem Analysis	0.8837	0.7842	1.000	0.7606	0.5926	0.6551	0.8967	0.6012
Delegation	0.8045	0.6718	0.7606	1.000	0.5506	0.4993	0.7064	0.6099
Interpersonal Skills	0.5823	0.6440	0.5927	0.5506	1.0000	0.3968	0.5616	0.4996
Stress Tolerance	0.6252	0.4972	0.6551	0.4993	0.3968	1.0000	0.5744	0.4241
Organization and Planning	0.8461	0.8709	0.8967	0.7064	0.5616	0.5744	1.0000	0.5172
Functional Knowledge	0.5943	0.5206	0.6012	0.6099	0.4996	0.4241	0.5172	1.0000

References

Asch, S. E. Forming impressions of personality. Journal of Abnormal and Social Psychology, 1946, 41, 258-290.

Bray, D. The assessment center and the study of lives. American Psychologist, 1982, 32 (2), 180-189.

Byham, W. C. The assessment center as an aid in management development. In K. N. Wexley & G. A. Yukl (Eds.), Organizational, behavioral, and industrial psychology. New York: Oxford University Press, Inc., 1975.

Byham, W. C. Assessment centers for spotting future managers. Harvard Business Review, 1970, 48, 150-160.

Byham, W. C., & Pentecost, R. The assessment center: Identifying tommorrow's managers. Personnel, 1970, 17-28.

Byham, W. C., & Wettengel, C. Assessment centers for supervisors and managers: An introduction and overview. Public Personnel Management, 1974, 3, 352-357.

Bynne, D. The repression-sensitization scale: Rational, reliability, and validity. Journal of Personality, 1961, 29, 334-349.

Buros, O. K. Personality tests and reviews II. New Jersey: The Gryphon Press, 1975.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 1977, 35 (1), 38-48.

Christie, R., & Geis, F. L. (Eds.) Studies in machiavellianism. New York: Academic Press, 1970.

Cooper, W. H. Ubiquitous halo. Psychological Bulletin, 1981, 90 (2), 218-244.

Cunningham, H. M. Assessment center performance of volunteers and nonvolunteers. Unpublished master's thesis, North Texas State University.

Darley, J. M., & Gazio, R. H. Expectancy confirmation processes arising in the social interaction sequence. American Psychologist, 1980, 35 (10), 867-881.

Eisen, S. V., & McArthur, L. Z. Evaluating and sentencing a defendent as a function of his salience and the perceiver's set. Personality and Social Psychology Bulletin, 1979, 5 (1), 48-52.

Firefighters Institutes for Racial Equality vs. City of St. Louis, U.S. Court of Appeals, Eighth Circuit (Appeal from the U.S. District Court for the Eastern District of Missouri) (Nos. 79-1435 and 79-1461) January 21, 1980.

Friend et al., vs. Leidinger et al., (City of Richmond) Civil Action No. 74-0327, U.S. District Court, Eastern District of Virginia, October 3, 1977.

Gough, H. G., & Heilbrum, A. B. The Adjective Checklist Manual. Palo Alto: Consulting Psychologists Press, Inc. 1965.

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

Huck, J. R., & Bray, K. W. Management assessment center evaluations and subsequent job performance of white and black females. Personnel Psychology, 1976, 29, 13-30.

Huguenard, T., Sager, E., & Ferguson, L. W. Interview time, interview set, and interview outcome. Perceptual and Motor Skills, 1970, 31, 831-836.

Jacobs, R., Kafry, D., & Zedeck, S. Expectations of behaviorally anchored ratings scales. Personnel Psychology, 1980, 33, 595-640.

Kelly, H. H. The warm-cold variable in first impressions of persons. Journal of Personality, 1950, 18, 431-439.

Kraut, A. I. A hard look at management assessment centers and their future. Personnel Journal, May 1972, 317-321.

Landy, F. J., & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87 (1), 72-107.

Leach, C. The importance of instructions in assessing sequential effects in impression formation. British Journal of Social and Clinical Psychology, 1974, 13, 151, 156.

Levenson, H. Distinctions within the concept of internal-external control: Development of a new scale. Proceedings of the 80th Annual Convention of the American Psychological Association, 1972, 7, 261-262. (Summary)

Mayo, C. W., & Crockett, W. H. Cognitive complexity and primacy-recency effects in impression formation. Journal of Abnormal and Social Psychology, 1964, 68 (3), 335-338.

Mensh, I. N., & Wishner, J. Asch on "forming impressions of personality:" Further evidence. Journal of Personality, 1947, 16, 188-191.

Moses, J. L. The development of an assessment center for the early identification of supervisory potential. Personnel Psychology, 1973, 26, 569-680.

Moses, J. L. Standards and ethical considerations for assessment center operations. Proceedings of the Third Annual Congress on the Assessment Center Method, May 1975, 1-10.

Nisbett, R. E., & Wilson, T. D. The halo effect: Evidence for unconscious alteration of judgements. Journal of Personality and Social Psychology, 1977, 35 (4), 250-256.

Rathus, S. A. A 30-item schedule for assessing assertive behavior. Behavior Therapy, 1973, 4, 398-406.

Slevin, D. P. The assessment center: Breakthrough in management appraisal and development. Personnel Journal, April 1972, 255-261.

Smith, A., & Cunningham, H. M. The new assessment center: An increase in content validity. Proceedings of the 28th Annual Convention of the Southwestern Psychological Association, 1982.

Snyder, M., & Swann, W. B. Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology, 1978, 36 (11), 12-2-1212.

Spence, J. T., & Helmrich, R. L. Masculinity and femininity: Their psychological dimensions, correlates, and antecedents. Austin: University of Texas Press, 1978.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. STAI manual. Palo Alto: Consulting Psychologists Press, Inc., 1970.

Srull, T. K., & Wyer, R. S. Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgements. Journal of Personality and Social Psychology, 1980, 38 (6), 841-856.

Srull, T. K., Wyer, R. S. The role of category accessibility in the interpretation of information about persons: Some determinants and implications. Journal of Personality and Social Psychology, 1979, 37 (10), 1660-1672.

Synder, M. Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 1974, 30, 526-537.

Task Force on Assessment Center Standards. Standards and ethical considerations for assessment center operations. Personnel Administrator, 1980, 25, 35-40.

Wilson, J. E., & Tatge, W. A. Assessment centers--further assessment needed? Personnel Journal, March 1973, 172-179.

Wollowick, H. B., & McNamara, W. J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 58 (5), 348-352.