Comparison of the Narratives vs. the Checklist Assessment Center Exercise Report Forms

1986

James C. Rehmann
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/rtd

University of Central Florida Libraries http://library.ucf.edu

STARS Citation

Rehmann, James C., "Comparison of the Narratives vs. the Checklist Assessment Center Exercise Report Forms" (1986). Retrospective Theses and Dissertations. 4866.
https://stars.library.ucf.edu/rtd/4866

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
COMPARISON OF THE NARRATIVE VS. THE CHECKLIST ASSESSMENT CENTER EXERCISE REPORT FORMS

BY

JAMES CARL REHMANN
B. S., Michigan State University, 1981

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Industrial/Organizational Psychology in the Graduate Studies Program of the College of Arts and Sciences University of Central Florida Orlando, Florida

Spring Term
1986
ACKNOWLEDGEMENTS

I would like to take this opportunity to thank those people without whose guidance and support, this achievement would not have been possible. I would first like to thank my thesis committee members, Dr. Wayne Burroughs, Dr. William Wooten, and Dr. Janet Turnage, for their guidance, concern, and exceptional flexibility in helping me meet impending deadlines.

I would also like to extend my sincere thanks to Michael Struth, who spent countless hours assisting me with the design and implementation of this study. His professional experience as well as sense of humor were essential ingredients for the successful completion of this paper. My thanks as well to Dave Bracken, and all of my other co-workers, who unselfishly offered their time and advice.

Finally, I want to thank my parents, Jack and Connie Rehmann, for their unceasing love and support. Despite my ever changing life, they have always been my Rock of Gibraltar. Thank you for always believing in me and teaching me that faith in our beliefs is really what we are all about.
# TABLE OF CONTENTS

## INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Assessment Center Method</td>
<td>1</td>
</tr>
<tr>
<td>How Assessment Centers Differ from Traditional Methods</td>
<td>4</td>
</tr>
<tr>
<td>Validity Support for Assessment Center Method</td>
<td>6</td>
</tr>
<tr>
<td>How Assessment Centers Can Be Improved</td>
<td>9</td>
</tr>
<tr>
<td>Research Objectives</td>
<td>10</td>
</tr>
</tbody>
</table>

## METHOD

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>14</td>
</tr>
<tr>
<td>Procedure</td>
<td>16</td>
</tr>
</tbody>
</table>

## RESULTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

## DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35</td>
</tr>
</tbody>
</table>

## APPENDICES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Check List Exercise Report Form</td>
<td>41</td>
</tr>
<tr>
<td>B. Narrative Exercise Report Form</td>
<td>51</td>
</tr>
<tr>
<td>C. Employee Discussion Criterion Exercise</td>
<td>56</td>
</tr>
</tbody>
</table>

## REFERENCES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>81</td>
</tr>
</tbody>
</table>
INTRODUCTION

Placing the right individual in the right job is probably the most traditional task of Industrial Psychologists. Improving this process of selection and placement of personnel has been one of the primary goals of Industrial Psychology since its inception. The criticality of this task has been even more magnified in recent times due to the increasing costs of technical training and the focus on quality of production. Organizations cannot afford large turnovers in personnel due to improper selection, nor afford to misplace personnel relative to their abilities and motivation.

Developing efficient and effective ways of hiring and placing personnel has led to a variety of methods. The most commonly used method is the employment interview. A survey conducted in 1958 indicated that 99% of the 852 organizations involved in the study used an interview before hiring applicants (Spriegel & James, 1958). It seems this popularity has withstood the test of time, although there are many dangers inherent in this method that threaten objective hiring and placement of individuals. Interviewers can be subject to biases due to personal prejudices, perceptions, and the relative differences in the applicant pool. Many of these pitfalls can be avoided though, by using trained interviewers in relatively structured interviews.
One of the biggest advantages of the employment interview is to provide information to the applicant about the organization at which they are applying. If interviewers are honest and frank about the job and the organization with the applicant, the predictability rate of selecting successful applicants has shown to increase significantly. The interview seems to have its greatest potential for assessing the motivation to work and the interpersonal competence of job applicants (Wexley & Yukl, 1977). It is for this reason that many selection systems incorporate an interview as part of their process of evaluation.

Typically functioning hand in hand with the employment interview is the job application form, which attempts to seek biographical data from the applicant. This method, too, can suffer the same pitfalls as the employment interview, although because of its inherent documentation qualities is more scrutinized by the Equal Employment Opportunity Commission (EEOC) and the Fair Employment Practice (FEP) laws. The EEOC and FEP laws make it illegal to ask certain questions pertaining to an applicant's race, religion, sex, national origin and age.

Another major problem with conventional application forms is the way in which they are developed. Often organizations use various versions of application forms employed at another organization, without first thoroughly investigating their own critical employment information needs. This leads to wasted time on the part of the applicant by filling out needlessly long application forms, and reflects a bad image on the organization as well.
If the application form is developed properly, such that it seeks only information critical for success on the job, it has been shown to be a useful selection and placement instrument. Several studies have shown its success, for example in predicting such criteria as tenure or turnover (Fleishman & Berniger, 1960), salary increase (Scollay, 1956), and performance ratings by supervisors (Scollay, 1957).

A large battery of psychological tests are available for selection and placement. Fourteen years previous to this writing, a publication that attempts to account for all published psychological tests listed over 1,200 on the market at that time (Buros, 1972). Today, the same publication reviews 1,409 tests (Mitchell, 1985).

With all these selection and placement options available it is necessary to establish methods of effectiveness. The degree of effectiveness, or validity, can be measured on many dimensions (e.g., face validity, content validity, etc.), but what most personnel administrators are concerned with is the predictive validity of an instrument. This is a measure of how well a particular test can predict candidate success on the job, or whether the selection device is capable of predicting subsequent behavior on the job.

Unfortunately psychological tests such as intelligence tests, and personality and interest tests may have fairly high validity for some jobs and no validity at all for others. It is unfortunate as well, that there are those who have the impression that these tests can predict success in almost all jobs.
Intelligence and ability tests can be useful tools only if used properly. They have been shown to be a better predictor of success in training than of actual job proficiency (Ghiselli, 1973). Bray and Grant (1966) found ability tests to be a fairly accurate predictor of salary progress.

Because intelligence and ability tests have shown inconsistent validity to various jobs, are generally costly to administer, and have met with close scrutiny by the EEOC regarding their content validity, it has led researchers to seek alternative methods of assessment. Many organizations still employ the use of such tests, but usually reserve their use to selecting high level managers.

**The Assessment Center Method**

Another alternative method that can be used for selection, placement, promotion or development is termed the assessment center. This term refers to a standardized off the job procedure used to identify and measure those skills of job encumbents that are necessary to be successful on the job. Although no two programs are alike, they all share the commonality of using multiple methods of assessment, multiple assessors, and ratings on several skill dimensions in simulated job situations (Thornton & Byham, p. 3).

Since their conception, assessment centers have met with controversy concerning their effectiveness and overall utility. The first industrial application of an assessment center, aptly named the Management Progress Study, was performed by the Michigan Bell
Telephone Company and AT&T (Bray, 1964). According to Bray, the study was instituted as a long-range research study of psychological development of adulthood. This study involved 355 newly appointed managers of AT&T. After eight years the center correctly identified some 80% of the original participants who eventually reached middle management. What it ultimately illustrated, and what had the most impact on the psychological and industrial world, was its identification and isolation of individual characteristics that lead to success as a manager. Bray (1964) had conceived an alternative to the traditional hiring and promotion methods that accurately predicted the actual progress individuals made in the company over the following years.

Bray's basic methodology set the groundwork for what is considered the traditional assessment center presently employed in thousands of organizations today. In this process, individuals participate in a series of situations that resemble what they might be called upon to do in the real world. These "real" simulations are intended to be situational examples extracted from the target position, or a position that requires filling. An example of this might be line workers participating in an assessment center to fill the position of first-line supervisor. The target position, the one the assessment center exercises are designed to simulate, is for the position of first-line supervisor. Thus, exercises in this assessment center would simulate actual tasks of the target position first-line supervisor in that organization. This offered a new approach to selection, promotion and development by evaluating
candidates not on what they have done in past or present jobs but on how they are likely to cope with a new type of position with different job responsibilities and degree of skill knowledge.

**How Assessment Centers Differ from Traditional Methods**

The rationale behind using such situational exercises is that they simulate the type of work to which the candidate will be exposed and allow his performance to be observed under somewhat realistic conditions. Contrary to the aptitude test approach, samples, not signs of behavior, are used for prediction (Thornton & Byham, 1982). One researcher feels these situational tests go beyond traditional methods of selection, such as interviews and traditional psychometric tests by measuring more complex or dynamic behavior rather than aptitudes or traits, for example: interpersonal skills, leadership, and judgement (Howard, 1971).

Situational methods offer the potential of adding to the scope of human characteristics which can be evaluated (Bray & Grant, 1966). Although the assessment center process is much more time consuming and expensive to administer than traditional forms of selection such as interviews and paper-and-pencil tests, the increased information on characteristics about the participant usually justifies the added costs.

Assessment centers have served many purposes. Taft (1959) points out that assessment centers have been used for personality research, selection, and validation of techniques. A present day
systems approach to assessment centers employs the process in selection, placement, training and development, and career counseling. Regardless of their intent, assessment centers focus on the observable. The design and intent of the centers are to evoke behaviors relevant to the job. Contrary to traditional methods of selection, very little, if any, emphasis is placed on projection of performance from various indices. This highlights another major difference between the assessment center process and traditional forms of assessment and selection.

Other unique characteristics of the assessment center approach are its use of multiple trained assessors and multiple exercises. The use of multiple assessors in a series of different simulation exercises allows for a number of advantages for objective rating. First, the candidate is required to perform in significantly different types of simulation exercises, all attempting to cover the gamut of behaviors required to perform the tasks in the target job successfully. Thus, many of the behaviors in the target position are represented, and differential performance in respect to those behaviors can be observed by the assessor. This also offers the candidate, for whatever reason, more than one shot at demonstrating his/her skills.

The use of multiple trained assessors adds to the reliability of the final assessment center ratings. Before any final ratings have been made, all the assessors that observed a specific candidate in the respective exercises meet to discuss exercise dimension ratings, and to decide overall final ratings. Although interrater
reliability is often high before this discussion takes place (Schmitt, 1977), this meeting requires all assessors to explain and often justify their rationale for their particular ratings.

The use of trained assessors, regardless of whether they are external or internal to the organization employing the assessment center, significantly reduces the effects of commonly found rating errors such as halo (tendency to rate a candidate high or low on all the measured skills or dimensions) or similar-to-me effects (rating the candidate relative to their similarity to the assessor) (Ivancevich, 1979; Thornton & Zorich, 1980; Latham, Wexley & Pursell, 1975). Often managers responsible for hiring and rating performance are not trained in the objective observation, recording, and rating of employee behavior. Thornton and Zorich (1980) have shown that as little as two hours' training can have a significant effect of reducing common rating errors and improve rating accuracy.

Rating employee or job candidate's performance is not necessarily unique to assessment centers, but rating the performance according to specific individual skills or dimensions was a novel approach. Bray and Grant (1966) first identified 25 characteristics of successful managers for use in their Management Progress Study, and later factor analysis of these rating variables yielded 11 factors for success. Many of these factors, e.g., interpersonal skills and administrative skills, are used in today's assessment centers. Rating on several variables has the advantage of demonstrating high and low performers on specific skills, and can
then be used for matching candidates to jobs that require varying degrees of certain skills for success on the job. It also offers the dual benefit of illustrating specific deficiencies in a candidate’s performance that can be used in tailoring individual or corporate training programs.

**Validity Support of the Assessment Center Method**

A review of the literature shows the validity of assessment centers as a predictor of managerial success as stable across different organizations and different managerial positions. In a review of predictive validities by Norton (1977), the literature on traditional methods for predicting managerial success reveals that the average validity of the assessment center is about as high as the maximum validity attained by use of traditional methods. One of the earliest, yet probably most important validity studies done on assessment centers is the AT&T Management Progress Study (Bray, 1964). The criteria variables of advancement and salary were uncontaminated because results of the center were not made known to the candidates of the organization. Point bi-serial correlation of an assessment center rating with actually making middle management was $r=.44$ for college graduates and $r=.71$ for non-graduates. In another review article by Cohen, Moses & Byham, (1974), results of 19 assessment center validation studies were summarized. The median correlation between assessment center performance and job performance was $r=.37$, median $r=.63$ in
predicting job potential, and median $r = .40$ in predicting job progress. They concluded that the assessment center is clearly a more valid method than other traditional methods of assessment in terms of subsequent rates of success of assessed and non-assessed groups.

Typically the success of the assessment process is subsequently validated against what happens in the organization under more naturalistic conditions, that is, whether or not in the normal course of events an individual is promoted. If this is used as the ultimate criterion against which an assessment program is validated, one might question why the process had been employed at all? The answer is that the assessment center program will be able to identify promotable people earlier in their careers, and it will help to clarify some of the skills important in promotion, and it may perhaps identify some people who should be promoted but who might under normal circumstances be overlooked. Also, when based on a thorough job analysis, content and predictive validity of the assessment center technique has been accepted by the EEOC (Henderson, 1979) and it is these same types of validity, predictive and content, on which the technique relies (Jaffee & Sefcik, 1980).

How Assessment Centers Can Be Improved

One of the drawbacks of the assessment center process probably cited most often are the costs incurred in the development and implementation of an assessment center. One study quotes the
costs of $45 to $1000 per participant (Millard & Pinsky, 1980). While it is probable that costs average around $500 per participant (Kolb, 1984), the costs per participant can incrementally reduce as the number of participants put through the center increases. Still, the development of a customized assessment center, from job analysis through development of materials, training of assessors and implementation of the center often precludes the use of the method by small organizations. Cost utility analyses for assessment centers in larger organizations overwhelmingly illustrate the benefits in this environment, where the opportunity exists for screening many individuals with one center.

Related to cost is the large time commitment required of the managers in an organization attempting to implement an assessment center. Time is often required of managers during the interview phase of the job analysis, in order to properly identify the critical skills necessary to be successful on the job. More often than not, managers within the organization are used as assessors during the actual implementation of the center, and thus require a training period of one to five days. During the actual assessment cycle, managers are then called away from their daily duties to act as assessors. The basic point here is that most managers' time is strictly limited to the daily tasks of managing, and large time commitments as those illustrated above can be burdening, at best, on the individual manager and organization alike.
This study is an attempt to reduce that time commitment required of writing exercise report forms, without losing any of the inherent validity of the assessment center process. Unlike other studies that aim at reducing the quantity of training (Thornton & Zorich, 1980; Latham, Wexley & Pursell, 1975) or reducing the size and number of exercises (Moses, 1973), or increasing the generalizability of the center (McConnel, 1971), while at the same time retaining high validity, this study looks at reducing the time spent by assessors on the exercise report form.

The exercise report form is traditionally a narrative type report that is filled out by an assessor following his or her observations of a participant in an interactive exercise, or in the case of a non-interactive exercise (e.g., in-basket exercise), during the review of the materials. One assessment center expert, (Struth, 1986), feels that a conservative estimate of the time required to fill out this form is between 45 to 90 minutes per exercise, depending on the type of exercise, and the degree of responses from the participant. Considering an average ratio of 1:3, assessors to participants, and that most assessment centers employ between four to seven exercises per cycle (Bender, 1973), it computes out to an average commitment on the part of an assessor of between 13.5 to 23.6 hours per cycle on exercise reports alone.

The narrative form, once completed usually consists of a synopsis of all the behaviors observed and recorded by the assessor during an exercise. The format for such a form can be several pages long, and on the top of each page is a definition of one of the skills
(e.g., leadership). Following this definition, the page is divided into two columns, one side reserved for positive behaviors concerning that skill, the other side for negative observations. This page format continues on for all the critical skills being used in that assessment center. The assessor’s role is to categorize all of his observations of the participant in the exercise and list them under the appropriate skill and column. Following this, an overall rating (often 1-7) is made by the assessor for each respective skill.

This study proposes a method designed to reduce the writing requirement in the exercise report writing phase of the assessment center process by offering an alternative exercise report form, a check list type report form. This type of exercise report form attempts to list all the relevant behaviors related to each of the respective skills measured in an exercise. In some instances, it admittedly may not be all inclusive of the range of behaviors falling under a particular skill, so space has been provided at the end of each skill's behavioral statements to write in those behaviors observed but not addressed in the form.

At the beginning of each skill's behavioral statements is the definition of that skill. Following that definition are behavioral statements relating to that skill. Beside each behavioral statement is a space for rating how the participant performed relevant to that statement. Immediately following each of the behavioral statements is a space provided to list any examples of skill-related behavior observed during the exercise. After all of the behavioral
statements have been rated, an overall rating is then made for each skill. This final numerical rating mimics the narrative form process.

By identifying and listing the critical behaviors required for successful performance beforehand, in respect to each of the critical skills, and allowing the assessors to rate the participant's performance regarding each of these behaviors on a continuum, it is hypothesized that this process/method is at least as valid, or more valid than the traditional free form, or narrative type exercise report form. The major benefit of this type of report form is that it takes far less time to complete compared to the traditional narrative form because of the reduced writing commitment. It may also be a more valid instrument for minimally trained and/or experienced assessors, because it ideally addresses all the behavioral information relevant to a particular skill, and forces a response to that effect.

Research Objectives

This study attempts to investigate comparable validity and inter-rater reliability, assessor preference, reduced completion time and accuracy of a check list exercise report form by comparing the independent ratings made by trained assessors on the traditional narrative form exercise report form to the structured check list type report form.
The following hypotheses were investigated in this study:

H$_1$: Assessors' mean skill ratings on the check list exercise report form compared to the mean ratings on the narrative exercise report form are not significantly different.

H$_2$: Inter-rater reliability of the assessor skill ratings on the check list exercise report form will be equal to, or higher than the inter-rater reliability of the assessor ratings on the narrative exercise report form.

H$_3$: Of the assessors that used the check list form, a significant number prefer its use over the narrative form.

H$_4$: Assessors rate the check list exercise report form as being, on the average, significantly faster to complete compared to the narrative form.

H$_5$: The check list exercise report form provides at least as good as, or significantly more accurate ratings of the behaviors vs. the narrative form, compared to an expert panel of assessors' ratings as criterion.
METHOD

Subjects

The subjects in this study were 33 undergraduate students fulfilling an optional research requirement for an Industrial/Organizational Psychology class. All the subjects were participating on a volunteer basis, and given extra credit in their Psychology course after completion of the study. Before participating, it was explained to the subjects that their participation was strictly voluntary, and that they could refuse to partake in any or all parts of the study at any time. Consent forms were then distributed and signed by all the subjects.

All the subjects were given three hours of assessor training before randomly being assigned to one of two groups. Each of the groups contained 17 individuals. The only criterion for participation was that the subjects had at no time received any formal training in the observation, categorization, and/or rating of human behavior.

Procedure

The group of 33 subjects was split, and two identical, three-hour training sessions were conducted for the purpose of providing thorough training to all participants. The focus of the
training was directed towards the instruction of objective observing, recording, categorizing, and rating of human behavior—specifically for an assessment center simulation exercise.

The training included instruction on how to avoid the primary pitfalls of objective observation, namely: halo effect, contrast effect, similarity and first impressions. Other areas of training familiarized the participants with the process of observing and recording behavior, then later categorizing and finally placing a numerical rating of one to seven on all of the three behavioral dimensions used in this study. The agenda of the training session is listed below, along with the time committed to each segment.

**Training Session**

<table>
<thead>
<tr>
<th>Duration</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 minutes</td>
<td>Overview of Activities</td>
</tr>
<tr>
<td>20 minutes</td>
<td>Note-taking Practice</td>
</tr>
<tr>
<td></td>
<td>- Do's and Don'ts (Emphasizing objective recording and specificity, etc.)</td>
</tr>
<tr>
<td></td>
<td>- Practice recording behaviors while viewing videotape</td>
</tr>
<tr>
<td>20 minutes</td>
<td>Introduction to the Skills</td>
</tr>
<tr>
<td></td>
<td>- Definition of skills and examples of behaviors that would fall under those skills</td>
</tr>
<tr>
<td></td>
<td>- Paper and pencil skill categorization exercise</td>
</tr>
<tr>
<td>15 minutes</td>
<td>Break</td>
</tr>
<tr>
<td>15 minutes</td>
<td>Review of Exercise and Assessor Guide</td>
</tr>
<tr>
<td>Time</td>
<td>Activity</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>20 minutes</td>
<td>Observe Videotaped Mock Candidate #1</td>
</tr>
<tr>
<td>10 minutes</td>
<td>Discuss Report Writing and Report Form</td>
</tr>
<tr>
<td>40 minutes</td>
<td>Write Up Report on Mock Candidate #1</td>
</tr>
<tr>
<td>20 minutes</td>
<td>Discuss Write Ups</td>
</tr>
<tr>
<td>10 minutes</td>
<td>Discuss Rating Process and Assign Rating</td>
</tr>
<tr>
<td>180 minutes</td>
<td></td>
</tr>
</tbody>
</table>

The three dimensions being employed in this study for the categorization and rating of behaviors are as follows:

1. **Leadership** - To direct and coordinate the activities of others; to delegate authority and responsibility; and to provide means of follow-up.

2. **Decisiveness** - To make decisions, render judgements, take action, independent of quality; and to defend decisions, judgements, and actions when challenged by others.

3. **Interpersonal** - To be sensitive and behave in ways which reflect the needs, feelings, and capabilities of others; to deal effectively with others regardless of status or position; to accept interpersonal differences and develop rapport with others.

As mentioned previously, assessor training included an orientation to these dimensions that incorporated behavioral examples of each.
A one-half hour videotape, simulating an actual role player and job incumbent, was presented to the participants in order to practice the observation, recording and rating process. A group discussion of the rating procedure followed this practice session in order to further orient the participants to the norms and expectancies of behaviors in simulation exercises.

All training was conducted by a professional in the human resource management field. He had several years experience in the private and public sectors, coordinating and developing simulation exercises as well as assessing and training assessors.

One week (seven days) after each respective group received its assessor training, it met again, whereupon the members viewed a videotape of another role player (Mock Candidate #2) acting in the same Employee Discussion exercise on which they were trained. They were asked to take comprehensive notes in order to rate this candidate as they had in practice sessions.

Immediately following the subjects' observing the videotape, they were randomly placed in one of two groups. One group was assigned a check list exercise report form designed specifically for this exercise (see Appendix A), and given 5 minutes of specific instructions about its content and use. The check list exercise report form developed for this study was designed to be a time-saving, comprehensive categorization of the behaviors expected to be observed in the Employee Discussion.
The check list started out with a definition of each dimension (e.g., leadership). Following this definition was a series of behavioral statements illustrating probable behaviors to be observed in the simulation. The assessor, or user of this form, was to simply check on a continuum of three minuses to three pluses, whether the behavior observed was highly negative to highly positive, respectively. Then the dimension was to be given an overall rating of one to seven, depending on the performance of the observed behaviors.

The second group was given instructions to use the traditional narrative type exercise report form, and given a review on its use. This form consisted of a definition of each dimension, and following were two blank columns—one illustrating an area for positive observed behaviors, and one for negative observed behaviors that fell under that dimension. An overall rating was then made for each dimension (see Appendix B).

Up to 35 minutes was allowed for both groups to complete their exercise report forms and rate the candidate's performance following the 15-minute videotape, then all the report forms were collected. After this period (the following week), the subjects were assembled again. At this time they were asked to fill out a short form concerning the length of time it took them to complete each respective form, as well as their preference of forms for assessment. Participants were then debriefed to explain the full purpose of this study, and allow for a question and answer period.
In order to assess the accuracy of assessor ratings for each respective form, an expert panel was formed. These experts all had a minimum of four days of intensive assessor training, and had participated in a minimum of two full assessment cycles. They viewed the candidate in the criterion exercise videotape, and rated him using the traditional narrative exercise report form.
RESULTS

Comparisons were made between forms on mean skill ratings and on inter-rater reliability of assessor skill ratings. Also, differences in the amount of time to complete each form, and assessor preferences of form were examined.

In testing hypothesis #1, a t-test (two-tailed) was used to determine differences between the assessor group means (Ferguson, 1981).

The means, standard deviations and t-test values comparing the narrative versus the checklist form group means on each skill are listed in Table 1. For t(31), an absolute value of 't' equal to 2.042 and 2.750 are required at the p<.05 and p<.01 significance levels, respectively. In this case, significant differences were found between the Leadership, and Decisiveness means for the two groups. There was no significant difference between the Interpersonal mean scores. From these results, it appears as though the check list form is in fact exerting some sort of differential effect on the assessment rating process for at least two of the three variables measured. Because there were significant differences between two of the three skills measured, there exist no grounds for accepting hypothesis #1.
TABLE 1

SKILL RATING MEANS, STANDARD DEVIATIONS, AND t-VALUES FOR GROUP SKILL MEANS BETWEEN CHECK LIST AND NARRATIVE GROUPS

<table>
<thead>
<tr>
<th></th>
<th>Check List Form</th>
<th></th>
<th>Narrative Form</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>3.375</td>
<td>2.000</td>
<td>5.813</td>
<td>4.588</td>
</tr>
<tr>
<td>STD. DEV.</td>
<td>1.025</td>
<td>.894</td>
<td>.981</td>
<td>1.372</td>
</tr>
<tr>
<td></td>
<td>Leadership</td>
<td>Decisiveness</td>
<td>Interpersonal</td>
<td></td>
</tr>
<tr>
<td>t-VALUES</td>
<td>-2.864**</td>
<td>-3.517**</td>
<td>.284</td>
<td></td>
</tr>
</tbody>
</table>

* p<.05
** p<.01
The inter-rater reliability (IRR) is an important concern when developing and employing a selection instrument. Hypothesis 2 investigates this concern by comparing the measurements of IRR for each skill between each exercise report form. As hypothesis #2 states, the check list form has at least as high an IRR coefficient as the narrative form.

In testing H₂, a method developed by James, Demaree & Wolf (1984) was designed especially for assessing agreement among the judgements made by a single group of judges on a single variable in regard to a single target. In this case, assessors' scores on each skill were compared between groups using the check list or narrative forms. The authors claim that their method of estimating IRR furnishes more accurate and interpretable estimates of agreement than estimates provided by procedures commonly used to estimate agreement, consistency, or inter-rater reliability. The method of obtaining an IRR estimate is obtained by placing the estimates of variances into the equation: \[(true\ variance)/(true\ variance + error\ variance)\]. The preceding term estimates the "proportion of random or error variance present in the observed ratings," and subtracting that sum from one "gives that proportion of non-error variance in the ratings, a reliability coefficient" (Finn, 1970).

The estimate of true variance is computed by dividing the sum of squared difference scores by one less than the total number of participants. The denominator in the IRR formula is the variance that would be expected if all the judgements were due exclusively
to random measurement, and is computed by squaring the number of possible responses, (in this case seven), subtracting one, and dividing by 12.

In order to test for significant differences between the inter-rater reliability estimates, those estimates were transformed into z-scores, and a test of significance was applied to those z-scores. To compare two z-scores, the process required subtracting one z-score from the other, and then dividing this remainder by the standard error of those z-scores (Edwards, 1984). The probability of this comparison Z-score was then determined from the table of the standard normal distribution. None of the differences between the IRR’s were found to be significant, yet simply eyeballing the data illustrates that the check list method produced a consistently higher inter-rater reliability compared to the narrative form for all skills measured. Table 2 illustrates the inter-rater reliabilities by skill for each form, as well as the comparative Z-scores.

Hypothesis #3 employs data from the opinion survey administered after the primary data were collected. In an attempt to assess what form the assessors preferred to use when assessing and rating, the assessors were asked what form they liked best, the check list or the narrative form. Although all subjects were trained using the narrative form, only half of the subjects used the check list report form when producing the criterion data, so only those individuals could participate in objectively reporting their preference of one form over another. Twelve subjects responded to
TABLE 2

WITHIN GROUP ASSESSOR INTER-RATER RELIABILITIES
BROKEN DOWN BY SKILLS
AND TESTS FOR SIGNIFICANCE OF DIFFERENCES IN RELIABILITIES

<table>
<thead>
<tr>
<th>Inter-Rater Reliabilities</th>
<th>Leadership</th>
<th>Decisiveness</th>
<th>Interpersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check List (N=17)</td>
<td>.7374</td>
<td>.8002</td>
<td>.7594</td>
</tr>
<tr>
<td>Narrative (N=16)</td>
<td>.5294</td>
<td>.3735</td>
<td>.6636</td>
</tr>
<tr>
<td>Expert Panel (N=9)</td>
<td>.8750</td>
<td>.9092</td>
<td>.8125</td>
</tr>
</tbody>
</table>

Comparison Z - scores

<table>
<thead>
<tr>
<th></th>
<th>Leadership</th>
<th>Decisiveness</th>
<th>Interpersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check List X Narrative</td>
<td>.671</td>
<td>.792</td>
<td>.122</td>
</tr>
<tr>
<td>Check List X Expert Panel</td>
<td>.082</td>
<td>1.310</td>
<td>.588</td>
</tr>
<tr>
<td>Expert Panel X Narrative</td>
<td>.446</td>
<td>1.910</td>
<td>.677</td>
</tr>
</tbody>
</table>

* p < .05
**p < .01
this item on the opinion/information survey, and in a tally of their responses it was found that all twelve (or 100%) of the assessors preferred the check list form, and none of the assessors (0%) responded that they preferred the narrative form. A chi square was performed on these data to determine the significance of these ratings, and a $X^2$ value as obtained with 1 df was 12.00, with $p<.0009$.

Hypothesis #4 attempts to determine whether the check list form is in fact faster to complete than the narrative exercise report form. Along with their form preference on the opinion/information survey, the subjects were questioned on how long it took them to complete their respective forms. The query allowed one of six responses along a continuum of 1 minute to 35 minutes. The response choices and frequencies for each form are listed in Table three.

In order to test for significance, the individual time ratings were averaged to allow for a test of mean times between the check list and narrative forms. For example, the third option on the the survey read: 16-20 minutes. This time span was averaged for analysis purposes to 18 minutes (see Table 3). Table 3 also shows the survey's means and standard deviations of reported completion times.

As can be seen in Table 3, the t-test method was employed to assess significance between the means. A $t(29)$ value of -2.286 was attained, which was significant at the $p<.05$ level. This indicates that the mean completion time for the check list was indeed significantly less than the mean completion time for the narrative form.
TABLE 3

REPORTED FREQUENCY OF FORM COMPLETION TIMES FOR THE CHECK LIST AND NARRATIVE GROUPS

<table>
<thead>
<tr>
<th>FREQUENCY</th>
<th>Check List</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 minutes</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(5.5 min.)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11-15 minutes</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>(13 min)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16-20 minutes</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>(18 min)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>21-25 minutes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(23 min)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>26-30 minutes</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>(28 min)</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>31-35 minutes</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(33 min)</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>

MEAN COMPLETION TIMES

<table>
<thead>
<tr>
<th></th>
<th>Check List</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.857</td>
<td>26.235</td>
<td></td>
</tr>
</tbody>
</table>

STANDARD DEVIATION

<table>
<thead>
<tr>
<th></th>
<th>Check List</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.419</td>
<td>6.660</td>
<td></td>
</tr>
</tbody>
</table>

t-Value

<table>
<thead>
<tr>
<th></th>
<th>Check List</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.286*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05
The premise behind H₅ is that the check list form is in fact more "accurate" than the narrative form, especially in this case, with minimally trained, inexperienced assessors. To establish a criterion for comparison, an expert panel consisting of nine well trained, experienced assessors viewed and rated the candidate in the criterion exercise videotape using the traditional, narrative exercise report form. This hypothesis was then tested by comparing the assessor ratings to those expert panel ratings.

The mean, standard deviation and index of inter-rater reliability, (IRR), was determined for the expert panel, employing the same method of determining IRR described for H₂, and are illustrated on Table 4. As was illustrated in Table 2, the expert panel and check list IRR ratings were not found to be significantly higher than the subjects using the narrative form, yet the expert ratings were somewhat higher than the IRR's for those subjects using the check list and narrative forms.

To test H₅ it was necessary to determine the degree of similarity of the check list and narrative forms to the expert panel's ratings. To do this, a comparison of difference scores of the subjects' ratings on the two forms, from the expert panel's mean scores, was made. The method required subtracting the panel of experts' mean rating on each of the skills, from each of the individual assessor/subject's ratings on that respective skill, and employing the use of the absolute value of that sum for analysis. The total, mean, and standard deviations for those difference scores for the narrative and check list forms are listed in Table 5.
TABLE 4

DESCRIPTIVE STATISTICS FOR EXPERT PANEL RATINGS

<table>
<thead>
<tr>
<th></th>
<th>Leadership</th>
<th>Decisiveness</th>
<th>Interpersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>3.333</td>
<td>1.889</td>
<td>5.333</td>
</tr>
<tr>
<td>STD. DEV.</td>
<td>.707</td>
<td>.601</td>
<td>.866</td>
</tr>
<tr>
<td>$R_{WG}$</td>
<td>.8750</td>
<td>.9097</td>
<td>.8125</td>
</tr>
</tbody>
</table>
# TABLE 5

**MEAN DIFFERENCE SCORES OF MEAN EXPERT PANEL RATINGS FROM INDIVIDUAL ASSESSOR SKILL RATINGS AND t-TEST VALUES COMPARING THOSE SCORES**

<table>
<thead>
<tr>
<th></th>
<th>Check List Form</th>
<th></th>
<th>Narrative Form</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>.792</td>
<td>.910</td>
<td>.903</td>
<td>1.530</td>
</tr>
<tr>
<td>STD.DEV.</td>
<td>.619</td>
<td>.364</td>
<td>.586</td>
<td>1.035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Check List</th>
<th>vs. Narrative</th>
<th>Leadership</th>
<th>Decisiveness</th>
<th>Interpersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-Value</td>
<td>-2.466*</td>
<td>-2.952**</td>
<td>-.681</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<.05

**p<.01
The difference scores were an indication of the similarity of each respective form's ratings to the ratings made by the expert panel. It was necessary, then, to determine if significant differences existed between these difference scores, and between the more accurate of the two forms, and the expert panel's ratings. A t-test was applied to the means between the difference scores, and finally between the mean skill scores of the subjects using the narrative and the check list, and the expert panel's ratings. The results of those t-tests are listed as well, in Table 5, and illustrates that for the skills of Leadership and Decisiveness the check list was significantly more accurate.

Further analysis was done to determine significant differences between the expert panel's mean ratings and those of the experimental group. The mean ratings were again examined at the skill level, and t-tests were used to compare significant differences between the expert panel, and the narrative and check list forms. The mean and standard deviations of the ratings from the narrative, check list and expert panel are listed for comparison in Table 6, as well as the results of the t-tests.

The results were similar to those attained when comparing difference scores. The check list's mean ratings were not significantly different from the expert panel's ratings for all three skills observed. The narrative form's ratings were found to be significantly different from the expert panel's ratings for two of the skills, Leadership and Decisiveness. There was no significant difference between the narrative form's mean rating and the expert
TABLE 6
CHECK LIST, NARRATIVE, & EXPERT PANEL MEAN SKILL RATINGS WITH STANDARD DEVIATIONS, AND RESULTS OF t-TESTS BETWEEN EXPERT PANEL MEAN SKILL RATINGS AND CHECK LIST AND NARRATIVE FORM RATINGS

<table>
<thead>
<tr>
<th></th>
<th>Check List</th>
<th></th>
<th></th>
<th>Narrative</th>
<th></th>
<th></th>
<th>Expert Panel</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lead</td>
<td>Dec</td>
<td>Int</td>
<td>Lead</td>
<td>Dec</td>
<td>Int</td>
<td>Lead</td>
<td>Dec</td>
<td>Int</td>
</tr>
<tr>
<td>MEAN</td>
<td>3.375</td>
<td>2.000</td>
<td>5.813</td>
<td>4.588</td>
<td>3.588</td>
<td>5.707</td>
<td>3.333</td>
<td>1.889</td>
<td>5.333</td>
</tr>
<tr>
<td>STD DEV</td>
<td>1.025</td>
<td>.894</td>
<td>.981</td>
<td>1.372</td>
<td>1.583</td>
<td>1.160</td>
<td>.707</td>
<td>.601</td>
<td>.866</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Check List Form vs. Expert Ratings</th>
<th>Narrative Form vs. Expert Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lead</td>
<td>Dec</td>
</tr>
<tr>
<td>t-Values</td>
<td>.108</td>
<td>.331</td>
</tr>
</tbody>
</table>

* p<.05
** p<.01
panel's mean rating for the skill of Interpersonal. The results support $H_5$, that the check list exercise report form provides at least as good, or more accurate ratings for all the skills when compared to the ratings of an expert panel as criterion.
DISCUSSION

In this study, a methodology was proposed to reduce the time commitment on the part of assessors when writing exercise report forms in assessment centers, without losing any of the inherent validity of the process. The study compared the proposed method to what is traditionally used and attempted to investigate comparable validity and inter-rater reliabilities, assessor preference, reduced completion time and the accuracy of the proposed method. The traditional method used for comparison in this study was termed a narrative exercise report form, and the proposed method was termed a check list exercise report form.

The results of this study generally confirmed the hypotheses stated previously in the Research Objectives section, with the exception of the first hypothesis. Hypothesis one postulated that there would be no significant differences in the measured skill ratings between the check list and narrative forms. The results did not fully support this hypothesis, because significant differences were found between two of the three skills employed in this study.

In interpreting these results, it is important to understand that the subjects in this study were undergraduate college students, who had only received three hours of assessor training, and previous to this study were generally unfamiliar with the assessment center process. Considering the amount of training and lack of experience,
relative differences in understanding the selected skills in this study could be expected. Significant differences between the respective form's mean skill ratings were found between the skills of Leadership and Decisiveness, and there was no significant difference between the mean ratings on Interpersonal. Perhaps the skills of Leadership and Decisiveness, and the behaviors that define those skills, were more difficult to comprehend for the subjects than for the skill of Interpersonal.

It can only be concluded at this point, that given the subject's amount of training and lack of assessor experience, the checklist exercise report form did exert a differential effect on the overall assessment center rating process, providing no support for hypothesis one. The logical question that presents itself following these findings, is given there are differences, what form is more accurate. This issue is addressed by hypothesis five.

The results supported hypothesis two, that the check list form would exhibit at least as high an inter-rater reliability, (IRR), as the narrative exercise report form. The IRR achieved by the check list form was somewhat higher, although not significantly higher than the IRR acquired with the narrative form. This increased IRR may be due to the explicit nature of the check list form, in that it induces increased consistency from the raters because it forces them to address a prespecified amount of skill relevant behaviors. The consistency between check list forms assists in increasing the consistency of what all the assessors rate on, and in turn, increasing their overall consistency, or reliability of ratings.
Hypothesis number three received unanimous support from the data, in that all the subjects that responded to the questionnaire item asking what form (check list or narrative) they preferred, chose the check list as their preference. It can be concluded then, that the check list form, for whatever reasons, is the overwhelming favorite of those subjects that used both types of forms.

One of the more critical hypotheses in regards to this study is number four, whether or not the check list form actually takes less time to complete than what is traditionally used. The results support this hypothesis by indicating more than a 20% average decrease in the amount of time it took the assessors to complete the check list form compared to the narrative form.

Although the decrease was found to be significant, it is possible an even greater decrease could be seen with experienced assessors. Inexperienced assessors typically spend a greater amount of time categorizing the behaviors and making overall skill ratings, simply due to their lack of familiarity with the process. Experienced assessors tend to spend a majority of their time writing the behaviors on to the exercise report form. Because the check list form attempts to reduce this writing requirement, use of this method by experienced assessors could illustrate an even greater time savings. This time savings would also compound by a factor of three or four, due to the fact that most assessment centers assess between 9 - 12 skills, and this study only addressed three. An additional savings per added skill could then be expected.
Hypothesis five attempts to illustrate the relative accuracy of the check list form. An inspection of the results indicates inconsistent results on the first analysis. Difference scores between the expert panel's mean skill ratings and those subjects using the respective forms, demonstrated the check list to be significantly closer to the expert's ratings for the skills of Leadership and Decisiveness. No significant differences were shown for the skill of Interpersonal.

Further analysis of comparing the group means with a t-test produced similar results; The check list was more accurate for Leadership and Decisiveness, and just as accurate as the narrative form for the skill of Interpersonal. Again, this could be a function of the skill of Interpersonal, and how it is understood by the subjects, relative to the other skills.

In summary, support was found for the hypotheses tested in this study, with the exception of hypothesis one. A differential effect in the rating process was demonstrated between the two types of forms. Further analysis of the data showed that the difference acquired was in fact a beneficial effect, for the check list was later shown to produce more accurate ratings when compared to an expert panel's ratings of the same criterion role play.

It should be emphasized that the current investigation employed the use of non-managerial individuals with a relatively small amount of assessor training and no experience, for subjects. The training and experience levels of the assessors in this study are typically different from what is actually found in assessment
center cycles. Although it could be argued that these subjects more closely resembled true managers in an actual real world situation, as compared to the professional assessors used to make up the expert panel, these subjects were in fact university students, which should be considered when drawing conclusions from the data in this study.

Further research should address the differential effects between skill ratings illustrated in this study. As suggested earlier, this could be a function of the subject's training and experience, and what may boil down to their understanding of the individual skills. Or, it could simply be a function of the skill, and what behaviors indicate a demonstration of that skill.

Another important point concerning the use of either form is the amount of behavioral information provided by each respective form. For purposes of writing final reports on the assessment center participants, and in order to give meaningful feedback to the participants regarding their performance, it is critical to have a maximum amount of behavioral information on the exercise report form. It may be found that the check list, in its present form, may be deficient in this area, and should require more behaviorally specific examples of the participant's performance. It is suggested that this weakness could be compensated for in the training of assessors, by encouraging the utilization of the space provided after each behavioral statement for writing in behaviorally specific examples (see Appendix A).
It is suggested that those responsible for the development and implementation of assessment centers consider the checklist method as a viable alternative to the narrative form that has traditionally been used, for the aforementioned reasons. Although the checklist requires more time in the developmental stages of an assessment center, its overall superior utility will be demonstrated in the implementation of the center by the time savings alone, and the other benefits that have been demonstrated in this study could be considered bonuses for trying something new.
APPENDIX A
CHECK LIST TYPE EXERCISE REPORT FORM
ASSESSOR CHECKLIST
FOR
COORDINATION EXERCISE

Candidate: ____________________________________________
Evaluator: ____________________________________________
Date: ____________________________________________

Use the following notation to indicate the degree of positive or negative performance exhibited by the candidate relative to all behavioral statements contained in the checklist based on what you observed him/her do ONLY IN THIS EXERCISE.

Extremely Positive Performance With
+++: Respect to the Behavioral Statement
Very Positive Performance With
**: Respect to the Behavioral Statement
Slightly Positive Performance With
+: Respect to the Behavioral Statement

Extremely Negative Performance With
---: Respect to the Behavioral Statement
Very Negative Performance With
--: Respect to the Behavioral Statement
Slightly Negative Performance With
-: Respect to the Behavioral Statement

0: Unobserved

Use the following rating scale to evaluate the overall performance exhibited by the candidate on each individual skill category measured by the exercise based on what you observed him/her do ONLY IN THIS EXERCISE.

7 - Outstanding
6 - Well Above Satisfactory
5 - Above Satisfactory
4 - Satisfactory
3 - Below Satisfactory
2 - Well Below Satisfactory
1 - Weak
0 - Not Observed

For any 100 candidates you might observe, the following distribution of ratings is likely to occur: 5 percent of the candidates are likely to be rated a “7”; 10 percent, a “6”; 20 percent a “5”; 30 percent a “4”; 20 percent a “3”; 10 percent, a “2”, and only 5 percent, a “1”. Remember, these percentages are by no means binding, and you may consider several candidates to perform in an outstanding manner on most skills; yet, when considering the entire group of candidates, the full range of skill levels should be observable.

The area immediately following each item is provided for the assessor to write in critical behaviors exhibited by the candidate that prompted the assessor to assign that ranking to that item. It is important for the assessor to write down the respective observed behaviors, especially if the ranking is extraordinary in any way.
Leadership: To direct and coordinate the activities of others; to delegate authority and responsibility; and to provide means of follow-up.

To what extent did the participant:

1. Initially take charge of the meeting by clearly stating the goals and purpose of the session. vs. failed to establish the purpose of the meeting or did so only late in the meeting.

2. Question and probe the employee for additional information and/or explanations for reported actions (e.g., “This is the information have...how do you see it?”). vs. failed to question the employee about his/her reported work deficiencies, etc.

3. Indicate what the specific problems were with the employee’s work habits (e.g., not filling out work tickets, poor work safety, respecting supervisors, and other work practices). vs. made no mention of the specific work problems.
LEADERSHIP-continued

To what extent did the participant:

_____ 4. Point out the consequences of specific problem behaviors to the employee (e.g., possible disciplinary action concerning careless work habits), vs. neglected to address the consequences of continuing to perform at a sub-standard level.

__________________________

__________________________

_____ 5. Clearly indicate what level of performance is expected in the future regarding the various performance problems (e.g., "You are to always wear rubber soled safety shoes when at work.") vs. made no indication as to what performance was to be expected.

__________________________

__________________________

_____ 6. Explain to the employee the reasons behind his/her decisions (e.g., "I need to transfer you to midnight shift because you’re the most experienced employee on the plate stretcher.") vs. made decisions without explaining the reasoning.

__________________________

__________________________

_____ 7. Maintain control of the session by not allowing the role players to digress vs. allowed the role players to digress or totally dictate what would be discussed.

__________________________
LEADERSHIP—continued

To what extent did the participant:

____ 8. Express his/her opinion (e.g., "I believe you should take more care in filling out work tickets.") vs. depended totally on the opinions and suggestions of the role player. ____________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

____ 9. Arrange for a specific monitoring and/or a follow-up meeting of some sorts, by which to assess the employee's subsequent performance vs. failed to establish any sort of follow-up in order to check on the employee's behavior. ______________

________________________________________________________________________

________________________________________________________________________

Additional Comments: ____________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________
DECISIVENESS

Decisiveness: To make decisions, render judgements, take action independent of quality; and to defend decisions, judgements, and actions when challenged by others.

To what extent did the participant:

1. Strongly express his/her opinion (e.g., "Yes, but...", "No, I disagree..."), vs. did not voice an opinion, or offered views/suggestions in a meek, apologetic manner? ___________

2. Demonstrate a minimal amount of hesitancy toward making decisions (i.e., about the drill, shift change, etc.) vs. avoided making decisions at all. ___________

3. Defend and maintain his/her position when challenged vs. allowed the role player to easily change his/her mind. ___________

4. Make specific decisions (e.g., injuries/accidents were caused by unsafe working practices of Johnson and not unsafe working conditions), vs. avoided making specific decisions. ___________
DECISIVENESS-continued

To what extent did the participant:

5. Force Johnson to defend his/her arguments vs. failed to question Johnson's argument/veiws. ____________________________
__________________________________________________________________________

6. Act willing to confront the role player that there are deficiencies in his work behaviors vs. neglected to address the specific work problems. ____________________________
__________________________________________________________________________
__________________________________________________________________________

Additional Comments: _______________________________________________________
__________________________________________________________________________
__________________________________________________________________________
INTERPERSONAL

**Interpersonal:** To be sensitive and behave in ways which reflect the needs, feelings, and capabilities of others; to deal effectively with others; regardless of status or position; to accept interpersonal differences and develop rapport with others.

To what extent did the participant:

1. **Immediately introduce himself/herself and seek to establish rapport prior to beginning a discussion of problem issues vs. immediately addressed the issues at hand.**

2. **Approach the session from a mutual problem solving approach vs. approached the session from a tough “Boss-Subordinate” position.**

3. **Acknowledge valid points made by the role player (e.g., “I see, so you don’t think it was your fault that you lost your drill?”), vs. failed to acknowledge valid points made by the role player.**

4. **Ask the participant about his side of the story vs. made conclusions and decisions based on the limited information available and ignored the role player’s input.**
To what extent did the participant:

5. Express confidence in the employee's ability to return to previous performance levels vs. cast doubt on the role player's ability to ever again perform satisfactorily.

6. Acknowledge the inconvenience to the employee posed by the shift change vs. matter of factly directed the role player to change shifts, disregarding the role player's schedule.

7. Explain his/her reasons for disagreement with the role player vs. offered no explanation to the role player when disagreeing with him/her.

8. Couch disagreement/criticism in a positive manner (e.g., "I want to address some issues concerning your work habits.") vs. used accusatory language and/or tone when disagreeing or criticizing work habits (e.g., "Your work has been very poor.").

9. Allow the role player to speak without interruption or used apologetic phrases when interrupting vs. interrupted the role player and neglected to use apologetic phrases when doing so.
APPENDIX B
NARRATIVE TYPE EXERCISE REPORT FORM
Situation: Participant is Pat Harris, a maintenance first line supervisor for Blue Chip Metal Works Company. He/she must conduct a counseling interview with Lou Johnson, whose performance has been causing some problems.

EXERCISE REPORT FORM

Participant: __________________ Date: __________________
Assessor: __________________ Exercise: Employee Discussion

Using the rating key provided below, rate the participant on each of the following skills based on what you have seen him/her do ONLY IN THIS EXERCISE.

Rating Key

7 - Outstanding 3 - Below Satisfactory
6 - Well Above Satisfactory 2 - Well Below Satisfactory
5 - Above Satisfactory 1 - Weak
4 - Satisfactory 0 - Not Observed

For any 100 participants you might observe, given that no prior screening has taken place with respect to the participants, the following distribution of ratings is likely to occur: 5 percent of the participants are likely to be rated a "7"; 10 percent, a "6"; 20 percent, a "5"; 30 percent, a "4"; 20 percent a "3"; 10 percent, a "2"; and only 5 percent, a "1." Remember, these percentages are by no means binding, and you may consider several participants to perform in an outstanding manner on most skills; yet, when considering the entire group of participants, the full range of skill levels should be observable (assuming that no prior screening has occurred).

Exercise Summary (for comments on unique or extenuating circumstances only)
Leadership: Rating ___

To direct and coordinate the activities of others; to delegate authority and responsibility and to provide means of follow-up.

Key Points:

- Controlled and guided the meeting with Lou.
- Delegated tasks to Lou to aid in solving the participant's performance problems.
- During the meeting with Lou, assigned specific tasks and then scheduled follow-up meeting to evaluate progress.

Positive | Negative

* These are general points; specific behaviors need to be listed
Decisiveness: Rating ___

To make decisions, render judgments, take action independent of quality and defend decisions, judgments, and actions when challenged by others.

Key Points*

- Confronted Lou with specific performance problems.
- Quickly ended irrelevant discussion and returned to the task.
- Demonstrated minimal hesitancy in making decisions.
- Responded to Lou's questions quickly and maintained his/her own positions when challenged by Lou.

| Positive | Negative |
Interpersonal: Rating _____

To be sensitive and behave in ways which reflect the needs, feelings, and capabilities of others; to deal effectively with others regardless of status or position; to accept interpersonal differences and develop rapport with others.

Key Points

• Attempted to establish rapport with Lou (greeted him/her; introduced himself/herself, etc.).
• Allowed Lou to speak without interrupting or interrupted in a polite manner ("Excuse me, if I may", etc.).
• Acknowledged Lou's positive past performance.
• Maintained a constructive, pleasant manner despite Lou's uncooperative attitude.

Positive | Negative
APPENDIX C

EMPLOYEE DISCUSSION CRITERION EXERCISE
BLUECHIP METAL WORKS
EMPLOYEE DISCUSSION EXERCISE
For the purpose of this exercise you are to consider yourself Pat Harris, a newly promoted maintenance first line supervisor in the Hot Line section of the Bartow Rolling Mill for the Bluechip Metal Works Company. Today's date is September 13.

Despite the fact that you are new to your position, you have been asked by your immediate supervisor, Jan Summers, to handle some personnel matters regarding one of your subordinates. The subordinate's name is Lou Johnson, a mechanic in the Hot Line section working on the swing shift which you supervise. Although you are not familiar with Lou Johnson or any of the matters which you will be discussing with Lou (this is your first time on the swing shift since you arrived), your supervisor has asked you to handle the matters in hopes that it will provide you with an opportunity to establish yourself with the craft personnel whom you supervise.

The matters which require coverage with Lou Johnson have been outlined for you in a memo from Jan Summers; the memo is enclosed in the materials which follow. This memo will more thoroughly explain exactly what your supervisor has requested of you. In addition to this memo, you will find an organizational chart, background information on Lou Johnson, etc. which will assist you in preparing for the meeting. You now have 20 minutes to review all of the materials contained in this packet and to prepare for your meeting with Lou. At the end of this time, you will then have 30 minutes to actually meet with and discuss the matters with Lou. At the end of the meeting, you will have 15 minutes to prepare a brief written summary covering your meeting with Lou.
Lou Johnson has worked in the maintenance department of the Bartow Rolling Mill for the last six years. Over the years Lou's performance has been satisfactory. In other words, Lou has been an average employee.

Although Lou has always gotten along well with fellow craft personnel, Lou's relationship with management and supervisory personnel has not been of the same quality. While not directly hostile toward supervisory personnel, Lou has been known to continually question the actions and decisions of supervision in light of the union contract. Relative to other employees, Lou is somewhat quick to file grievances over even small matters that are, to most workers, unimportant. Although not really viewed by supervision as a "troublemaker," Lou has earned the reputation of being somewhat difficult to supervise.

Lately, Lou's uncooperative behavior appears to have become somewhat more pronounced. Recently, some supervisors have complained that Lou has become more aggressive toward supervisors, and on a few occasions Lou's comments have bordered on insubordination. A factor which could have some bearing on Lou's behavioral change may be personal problems Lou has experienced. Lou is currently involved in a divorce suit. This may or may not account for Lou's more aggressive manner toward supervision.
Pat Harris  
Hot Line Maintenance  
First Line Supervisor

HYDRAULICS SPECIALIST:  
Carl Evans

MECHANICS  
Lou Johnson  
Scott Beal  
George Sands  
Richard Jackson  
Sandy White  
Fred Green  
William Lawson  
Buddy Hicks

PIPEFITTERS  
Hank Fine  
Ted Manning

WELDERS  
Art Kline
BLUESHIP METAL WORKS

MEMORANDUM

TO: Pat Harris, Supervisor
FROM: Jan Summers, Manager
DATE: September 13
RE: Lou Johnson

Let me again say "welcome aboard" and at the same time apologize for putting you on the spot with this Lou Johnson matter; however, I feel that giving you the opportunity to act on problems like this right from the onset will go a long way toward your establishing yourself with the craft personnel.

There are several matters which I feel you need to discuss with Lou and I have outlined them below.

GENERAL PERFORMANCE: Over the last few months, Lou's work performance has dropped off a good bit, and I think we need to take some action to make sure that it doesn't get out of hand. Although Lou's never been what I would call "outstanding," he's usually been pretty reliable, but lately there have been some problems.

Over the past few months Lou has developed "sloppy" work habits. I've received several complaints from production regarding slip-shod jobs that have cost them time. Also, Lou has had some careless accidents that nearly caused serious damage to products and endangered himself and co-workers. I've included some material to specifically
I think some firm discussion on these problems will help Lou to shape up to get more in line with the satisfactory performance of which he is capable. With the pressure we are feeling from production, we can't afford any letdown by any of the craft personnel. Thus, I think some firm counseling with Lou is required.

Lately there have been some minor problems with the plate stretcher. Although the problems have not yet presented serious operational difficulties we need to get that piece of equipment in top working condition very quickly. Next week we begin production on the Simmons Shipyard contract, and we will, of course, be using the stretcher on almost a continuous basis during that production period. The contract has some time constraints, so we really can't afford any excess down time on the stretcher.

To correct the problems with the stretcher, I'd like Lou to move over to the midnight shift for all of next week. During the midnight shift next week, we plan to iron out the minor problems and give the stretcher a thorough trouble-shooting. Lou has more mechanical experience on that piece of equipment than does anyone else. Lou was very involved in its installation six years ago and has been the key mechanical craft person to service it. This is why we need Lou to switch to the midnight shift for all of next week. Although the company typically doesn't like to move people across shifts like this, I checked with industrial relations, and they said it was not a problem so far as the union contract was concerned, as long as we could show a reasonable need to take such action.
Temporary Shift Change: Continued

I've arranged for Lou to come by and speak with you today regarding some lost tools. Lou approached me about replacement of these lost tools, but I told him to take the matter up with you. As you know we have a strict rule that requires lost tools to be replaced by the employee, except under extenuating circumstances deemed reasonable by the supervisor. As far as Lou's concerned, that's the only purpose for the meeting between the two of you.

Again I apologize for putting all of this on you so soon, but I'm sure you will handle the matter appropriately. After you have finished the meeting with Lou, I'd appreciate a brief narrative summary of the meeting for my files, covering what was discussed and what you feel you accomplished with Lou.
To: P. Harris
From: Lee Johnson
Date: September 4
Re: Replacement of Tools

Yesterday while I was working in the casting area, the production supervisor accidentally kicked my drill off the platform I was working on and it fell into a pouring mold. Needless to say, I now don't have a drill.

I don't think that I should have to pay for the drill since I wasn't responsible for the loss. I told Mr. Summers about this, but he said I'd have to see you. I'd like to see you about this as soon as possible.

Sincerely,
I have received several complaints over the last few months from various production supervisors and managers regarding the quality of work performed by your work group. I feel that we have a problem that needs attention. Although the complaints vary, they share one common element - the involvement of Lou Johnson.

It seems that Johnson has on several occasions failed to complete work on equipment in a thorough manner, and this has resulted in equipment having to be removed from production a second, and sometime a third time; a thorough repair job on the first effort would have alleviated the need for additional service and reduced down time. I have listed below some incidents which I feel are representative of the types of complaints I am referring to.

1. Several weeks ago the overhead crane broke a drive belt, and Johnson came over to put a new one on it. Two days after this, the belt broke again. Johnson again replaced the belt. The next day, the same thing happened. This time, Fred Green came over and replaced the belt. In addition, he found that it was necessary to do some minor realigning; the slight misalignment was causing the belt to wear excessively on one edge. The wear was just enough to cause the belt to break in just a short period of use. Had Johnson taken the time to check the alignment the first time the belt broke, the additional work and down time would have been avoided.
2. Several months ago there were some problems with the plate cutter, an oil leak that we couldn't trace. Johnson looked it over and found a leaking hose. He replaced this hose only to find that the leak was not stopped. Upon a closer inspection, Johnson determined that a faulty hose clamp was the cause of the leak, and he replaced the clamp. This did, in fact, stop the leak. The problem is that what could have been a ten-minute job became an hour-long job because Johnson did not take the extra time to determine what the exact cause of the leak was before starting repairs.

There have been other incidences similar to these, and some of the production supervisors and managers are complaining. I would appreciate your taking some corrective action on this matter as soon as possible. Down time due to mistakes such as these really cost my people time, and it becomes more and more difficult for them to meet production quotas.
TO: Jan Summers, Manager
FROM: Jack Elliott, Superintendent Administration
DATE: September 10
RE: Incomplete Work Tickets

As you know, an accurate, detailed and thorough description of work completed by the craft personnel is essential in maintaining historical records for assistance in our maintenance planning functions. Lately, my office has experienced increasing problems with job tickets completed by craft personnel. The descriptions of the work performed are not detailed nor thorough in some instances, and this results in my people having to locate the employee who performed the work in order to obtain the missing information. I'm sure you realize that this means the needless loss of time for both your people and mine.

I'd very much appreciate your discussing this with your people in general. I think that you should take this matter up with several employees in particular, as they seem to be some of the prime offenders. I have listed them below.

1. Lou Johnson
2. Carl Evans
3. Zack Tailor
4. Wilson Thompkins

Thank you very much for your cooperation in this matter.
Employee: Lou Johnson

Reporting Period: January 1 (this year) to present

Description of Injuries/Accidents in Order of Occurrence

1) January 22:

While climbing the stairwell leading to the operator's room of overhead crane #4 in section I, employee fell and bruised left hip. This accident resulted in the employee's being unable to report to work for the following two days; however, the accident was not severe enough to require outside medical treatment.

Cause of the accident was claimed by the employee to be unsafe working condition, specifically excessive grease and oil on the steps of the stairwell. However, the conclusion of the safety committee was that the cause of the accident was unsafe work practices of the employee. When the accident occurred, the employee was wearing leather-soled shoes instead of rubber soled safety shoes as advised in section 19.1 of the company safety manual.

2) July 23:

While attempting to adjust the safety screen on a portable electric heater in the employee's lunchroom, the employee received minor burns on the thumb and index finger of the right hand. This accident resulted in no lost work days for the employee, and only minor first aid was administered at the company first aid clinic.

The conclusion of the safety committee was that the cause of the accident was unsafe work practices by the employee. The employee attempted to adjust the safety screen of the unit while the unit was in operation.
3) August 20:

While attempting to hand saw a piece of galvanized steel pipe, the employee received minor cuts and abrasions on the thigh area of the left leg. The accident resulted in no lost work days for the employee, and only minor first aid was administered at the company first-aid clinic.

The conclusion of the safety committee was that the cause of the accident was unsafe work practices by the employee. The employee attempted to saw the pipe in a careless manner rather than take the pipe to a work area where proper equipment and tools were available to perform the task in a safe manner.
TO: Jan Summers  
FROM: Bill October, Production Supervisor - Casting  
DATE: September 7  
RE: Lou Johnson

I really think that you better have a talk with Lou Johnson in the near future about the way Lou acts toward other people.

The other day Lou was over here in casting adjusting the clutch on the overhead crane so that we could go ahead and complete some of the last pourings required on an order before the end of the shifts. I walked over and asked Lou how much longer it would take to finish adjusting the clutch, and he jumped down my throat. Lou said, "If you need it any faster than I can fix it, maybe you ought to try fixing it yourself."

I realize that everyone, your people and my people, are under a lot of pressure due to production quotas, but I can't let things like that happen. If something like this happens again, I'll have no choice but to write Lou up.
BLUECHIP METAL WORKS
EMPLOYEE DISCUSSION EXERCISE
Assessor Guide

This Guide Contains:
• Administrative Instructions
• General Role Play Instructions - Lou Johnson
• Role Responses to Specific Issues - Lou Johnson
• Behavioral Skills
ADMINISTRATOR INSTRUCTIONS

1. Distribute the exercise materials to the participant and inform him/her that he/she has 20 minutes to review the material.

2. At the end of the 20-minute review period, usher in the role player and inform the participant that he/she has 30 minutes to meet with Lou Johnson.

3. After the 30 minutes allotted for the meeting, inform the participant that he/she now has 15 minutes to prepare a written summary of the meeting with Johnson.

4. At the end of the 15-minute period, collect all materials from the participant.
GENERAL ROLE PLAY INSTRUCTIONS

-Lou Johnson-

For the purpose of this exercise you are to assume the role of Lou Johnson, a mechanic in the Hot Line section.

Today, you are scheduled to meet with your supervisor, Pat Harris (the participant). You are under the impression that the sole purpose of this meeting is to discuss the replacement of a lost drill. When issues regarding problem performance are raised, you are to become defensive. Seek to minimize and rationalize the problems.

You do not feel that your performance has declined. In some instances, you are to take a strong exception to the accuracy of the information. In other instances, you are to state that the problems are a function of situational variables (heavy maintenance demands) which are beyond your control.

If the participant does not agree with the excuses you give, become somewhat angry. You should state that the participant's position is unfair. You are to maintain the attitude throughout the session.
SPECIFIC RESPONSES TO SPECIFIC ISSUES
-Lou Johnson-

- REPLACEMENT OF THE LOST DRILL (ITEM 2):

You are adamant in the belief that you should not have to pay for the replacement of the drill. If the participant indicates that you will have to pay for the drill, become hostile and say, "There's no way I'm paying for the drill. I'll file a grievance on this mess, but I'm not paying for the drill." Argue the point strongly until the participant specifically indicates that the matter is no longer open for discussion during the session.

- PERFORMANCE PROBLEMS:

1. Lack of thoroughness on initial work efforts which necessitates rework (Items 3, 3a).

   If the participant mentions the issue in a broad or general manner, become very defensive and demand that the participant be specific in his/her "criticism." When the participant becomes more specific, act amused and say: "You have to be kidding!" Then become very defensive and state that with the mill running flat out, maintenance demands are heavy, and everyone is "fighting brush fires." Reason that you do the best you can and then conclude your argument by saying: "I guess when you don't actually have to do the work, it's always easy to talk about how it ought to be done!" (In a very hostile tone) you maintain this line of reasoning on this issue. Other comments you can make on this issue are:

   1) "If it's so easy out on the floor, why don't you just come out and show us how."

   2) "I hope you're planning to call everyone else in here and give them the same speech because everyone here has the same problem, and it's not because we don't know how to do our job. There's too much to do."

2. Lack of thoroughness on job tickets (Item 4).

   You feel that this is a very minor complaint and don't understand why administration is so "picky" and wants "every little detail" regarding a repair. If the participant pushes on this issue, making comments like: "What the heck do you want us to do, fill out paper or get the
Specific Responses to Specific Issues - Continued

2. Lack of thoroughness on job tickets (Item 4) continued:

"If it's not one thing it's another. I can waste half the time fooling with paper if that's what you really want!" Also be sure to specifically make the comment, "How am I supposed to know what they want on those things? I just put down what seems right." Your point here is to subtly indicate that you, in fact, do not understand exactly the information required on the work tickets.

3. Safety Report (Items 5, 5a)

If the participant makes broad or general reference to safety issues, demands more specificity by saying: "What do you mean? I've never had any safety problems." After the participant becomes more specific say, "First of all, I don't think I'd call a few minor cuts and bruises a safety problem. Secondly, why are you walking in here as a new supervisor and bringing up things that happened before you got here? Is that the way you want things to be, boss?"

4. Conflict with Supervisor (Item 6)

First deny that the incident happened. If probed, admit that the incident occurred but that it was not your fault. Say that the supervisor was "trying to rush me when I was already working as fast as I could." If pressed, continue to put the blame on the production supervisor, indicating that production personnel are under stress due to production quotas. You can add, "I don't have to take that kind of stuff off of anyone," (meaning the production supervisor). At no time admit that you are in anyway to blame for the incident.

TEMPORARY SHIFT CHANGE:

You do not want this assignment due to some previous social commitments (i.e., some late night plans with friends on a couple of nights next week). If the participant asks you to take the assignment after discussing your performance problems, say: "You know, you really have a lot of nerve. First you rake me over the coals and then you want me to help you out. No way!" If the participant pushes for you to accept the assignment, ask why someone else can't be switched. If
Specific Responses to Specific Issues - Continued

pushed further, say, "I don't think this is in line with the union contract." You neither want nor do you willingly accept the assignment; however, if the participant makes it clear that you are being assigned, you have no choice but to do the job.

- **PERSONAL PROBLEMS (DIVORCE)**

  If asked generally or specifically about any personal problems, state that these matters are in no way affecting your work. You will not discuss the matter in detail.

- **TANGENTIAL DISCUSSION**

  At an appropriate point in the discussion, somewhere between the 15- and 20-minute point, you are to attempt to engage the participant in a tangential discussion. The discussion should begin with comments regarding the demands on the mill for maximum productivity and then quickly move into predictions of higher future markets and possible general expansion of the Rolling Mills to keep pace with market demands. The point here is to see if the participant will bring the discussion back to the main topics. You continue the discussion of the tangent for no more than three minutes or until the participant clearly indicates that the discussion needs to move back to central issues.
Leadership: To direct and coordinate the activities of others; to delegate authority and responsibility; and to provide means of follow-up.

1. Did the participant control the course of the meeting rather than allow the employee to determine the scope and topics of discussion?

2. Did the participant direct the discussion back to central topics when the employee attempted to digress?

3. Did the participant question and probe the employee for additional information and/or explanations for reported actions?

4. Did the participant offer his/her own views on the issues to the employee?

5. Did the participant clearly indicate what level of performance is expected in the future regarding the various performance problems?

6. Did the participant explain the reasons behind his/her conclusions to the employee?

7. Did the participant point out the consequences of specific problem behaviors to the employee?

8. Did the participant arrange for a follow-up meeting of some sorts, at a later time, by which to assess the employee's subsequent performance?

9. Did the participant clarify the aims and goals of the meeting with Johnson?
Behavioral Skills - Continued

Decisiveness: To make decisions, render judgments, take action independent of quality; and to defend decisions, judgments, and actions when challenged by others.

1. Did the participant make specific decisions (i.e., injuries/accidents were caused by unsafe working practices of Johnson and not unsafe working conditions)?

2. Did the participant maintain his/her position throughout the meeting with Johnson?

3. Did the participant demonstrate a minimal amount of hesitancy toward making decisions?

4. Did the participant defend his/her position when challenged (i.e., that Johnson's work habits were becoming sloppy)?

5. Did the participant strongly express his/her opinion (e.g., "Yes, but...", "No, I disagree...")?

6. Did the participant force Johnson to defend his/her arguments?

7. Did the participant make a decision when called upon to do so?
Behavioral Skills - Continued

Interpersonal: To be sensitive and behave in ways which reflect the needs, feelings, and capabilities of others; to deal effectively with others regardless of status or position; to accept interpersonal differences and develop rapport with others.

1. Did the participant initially seek to establish rapport prior to beginning a discussion of problem issues?
2. Did the participant approach the session from a mutual problem-solving approach as opposed to approaching the session from a tough "Boss-Subordinate" position?
3. Did the participant acknowledge valid points raised by the employee?
4. Did the participant acknowledge the employee's positive intentions regarding the accident with saw (i.e., trying to complete the job quickly)?
5. Did the participant ask the employee for "his/her side of the story" on issues, opposed to assuming that the employee is "guilty"?
6. Did the participant express confidence in the employee's ability to return to previous performance levels?
7. Did the participant acknowledge the inconvenience to the employee posed by the shift change?
8. Did the participant remain polite throughout the exercise (e.g., used please, thank you, etc.)?
9. Did the participant continue to interact with Johnson in a pleasant manner when conflict occurred, as opposed to withdrawing or becoming angry?
10. Did the participant explain reasons for disagreement with others?
REFERENCES


